

# Are Neural Click Models Pointwise IPS Rankers?

Philipp Hager<sup>1</sup> Maarten de Rijke<sup>1</sup> Onno Zoeter<sup>2</sup>

<sup>1</sup>University of Amsterdam <sup>2</sup>Booking.com

## Introduction

Inverse-propensity scoring (IPS) and click models (CM) are prevalent methods for learning rankers from position-biased click data:

- **IPS methods:** Assume position bias is known, typically use document features and optimize point-/pair-/listwise rankings.
- **Click models:** Infer position bias, typically treat each document separately and make pointwise relevance estimations.

Is a neural click model leveraging document features equivalent to a pointwise IPS ranker when both use the same position bias estimation?

- We compare both approaches theoretically.
- We perform an empirical comparison on semi-synthetic click data.

## Methods

**Neural click model:** Predicts biased user clicks. Document relevance  $\hat{y}_d$  and position bias  $\hat{o}_k$  are latent parameters inferred by minimizing binary cross-entropy between predicted and observed clicks [4]:

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

**Pointwise IPS:** Directly predicts document relevance  $\hat{y}_d$  by weighting clicks inversely to the probability of being observed by the user [3]:

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

## Comparing unbiasedness

- **IPS:** Saito et al. show that the pointwise IPS estimator is unbiased when correctly estimating position bias [3].
- **CM with joint parameter inference:** Oosterhuis shows that click models jointly inferring bias and relevance are not always consistent estimators of document relevance [2].
- **CM only inferring relevance:** We show that neural click models optimize for unbiased document relevance when, similar to IPS, given access to the true position bias:  $\hat{y} = \frac{oy}{\hat{o}}$ .

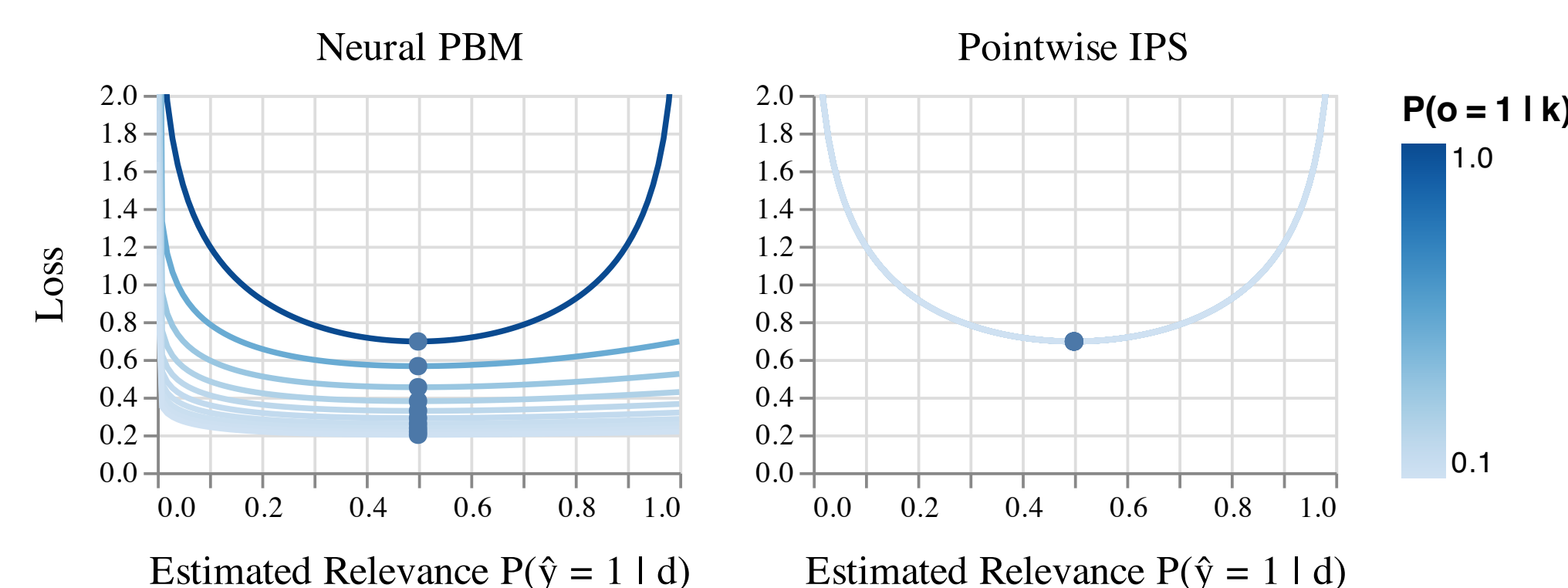


Figure 1: The loss for a document of relevance  $y_d = 0.5$  under varying degrees of position bias. Both methods have their minima at the true relevance, but the magnitude of the click model loss decreases with an increase in bias.

## Experimental results

**RQ:** Given the true position bias, do both approaches perform equally in an empirical comparison?

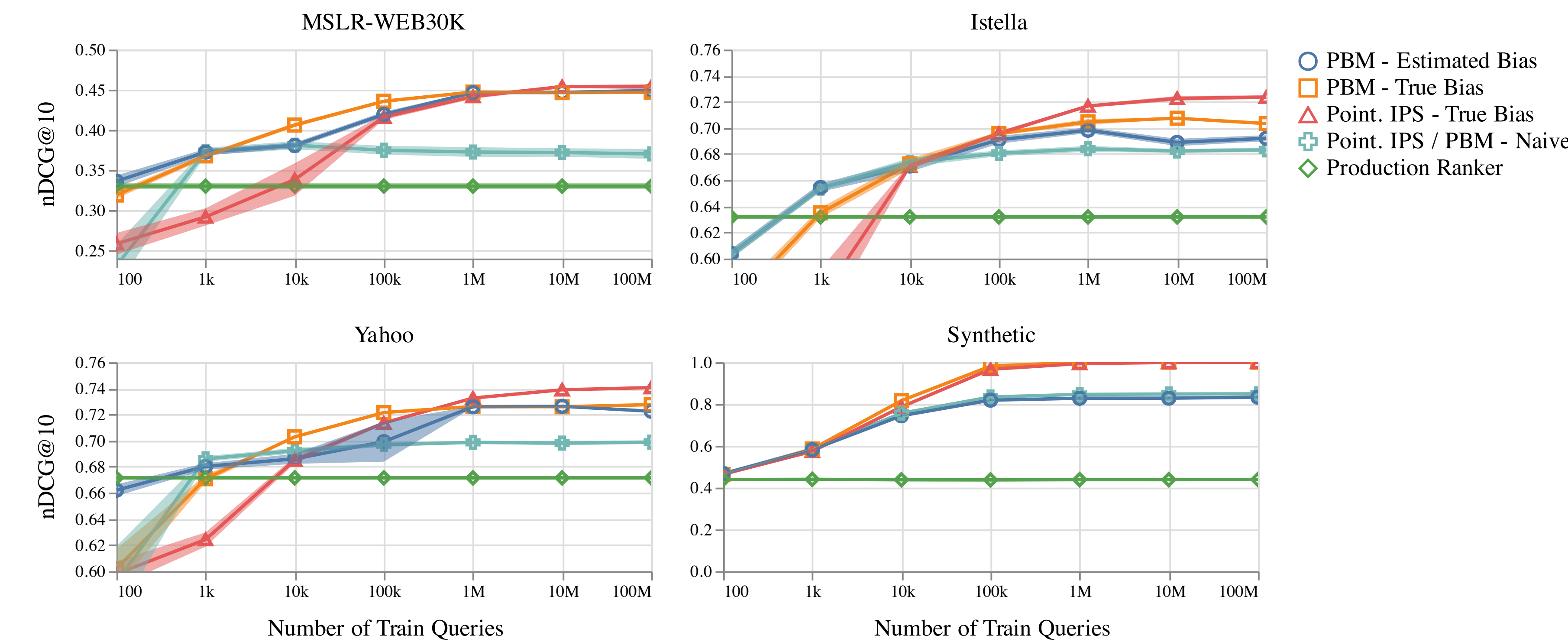


Figure 2: Test performance on three large-scale LTR datasets and one fully synthetic dataset after training on up to 100M simulated queries. The simulated user behavior follows the position-based model [1]. Results are averaged over ten independent experimental runs.

- **PBM - Estimated Bias:** A neural click model jointly estimating position bias and relevance performs better than a naive baseline that does not compensate for position bias on Yahoo and MSLR-WEB30K.
- **PBM - True Bias:** A neural click model with access to the true position bias has reduced variance and improved performance over the naive baseline on all datasets.
- **IPS - True Bias:** Pointwise IPS outperforms the neural click model significantly on Istella and Yahoo.

## Is the neural click model biased?

**RQ:** Items at lower position contribute less to the click model's loss, does generalizing over features introduce bias?

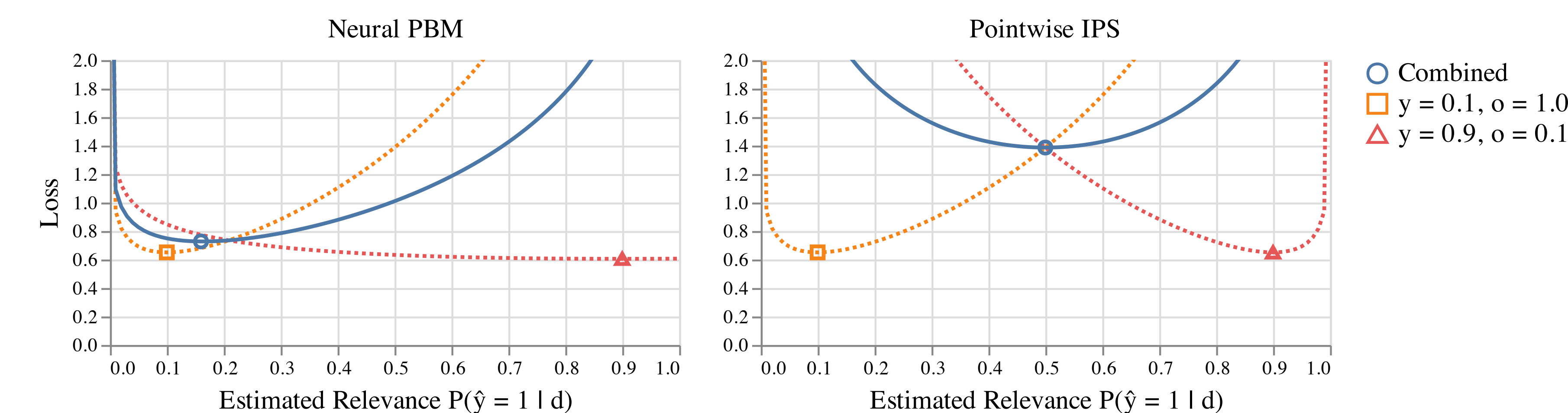


Figure 3: Visualizing loss and estimated relevance for an irrelevant document displayed at a high rank (orange square) and a relevant document displayed at a low rank (red triangle). The solid blue line shows the combined loss of both documents.

- Both approaches converge to the unbiased document relevance when computing the loss for each item separately (dotted lines).
- IPS converges to the average relevance of both documents when optimizing the combined loss for both documents (solid line).
- The neural click model converges to a joint document relevance for both documents that is biased towards the item with the higher examination probability.

## Experiments on neural click model bias

We run three additional experiments:

- **Independent features:** When documents share no features (one-hot encoded documents), the empirical performance of IPS and the neural click model is equivalent (Synthetic in Figure 2).
- **Feature collisions:** Gradually forcing random documents to share feature vectors leads to a stronger performance drop for the neural click model.
- **Position bias recovery:** When simulating increasing levels of position bias, we find that IPS can recover from strong position bias, while the neural click model increasingly deteriorates in performance even if the true position bias is known.

## Conclusion

- We show that both approaches optimize for unbiased document relevance if the true position bias is known and relevance is estimated separately per query-document pair.
- We find the neural click model to be affected by position bias when learning from shared, sometimes conflicting, features instead of estimating each document's relevance separately.

## Selected references

- [1] CRASWELL, N., ZOETER, O., TAYLOR, M., AND RAMSEY, B. An experimental comparison of click position-bias models. In *WSDM '08*.
- [2] OOSTERHUIS, H. Reaching the end of unbiasedness: Uncovering implicit limitations of click-based learning to rank. In *ICTIR '22*.
- [3] SAITO, Y., YAGINUMA, S., NISHINO, Y., SAKATA, H., AND NAKATA, K. Unbiased recommender learning from missing-not-at-random implicit feedback. In *WSDM '20*.
- [4] YAN, L., QIN, Z., ZHUANG, H., WANG, X., BENDERSKY, M., AND NAJORK, M. Revisiting two-tower models for unbiased learning to rank. In *SIGIR '22*.



Scan to read the paper



UNIVERSITY OF AMSTERDAM