



Unbiased Top- k Learning to Rank with Causal Likelihood Decomposition

Haiyuan Zhao
School of Information
Renmin University of China
Beijing, China
haiyuanzhao@ruc.edu.cn

Jun Xu*
Gaoling School of Artificial
Intelligence
Renmin University of China
Beijing, China
junxu@ruc.edu.cn

Xiao Zhang
Gaoling School of Artificial
Intelligence
Renmin University of China
Beijing, China
zhangx89@ruc.edu.cn

Guohao Cai
Noah's Ark Lab, Huawei
Shenzhen, China
caiguohao1@huawei.com

Zhenhua Dong
Noah's Ark Lab, Huawei
Shenzhen, China
dongzhenhua@huawei.com

Ji-Rong Wen
Gaoling School of Artificial
Intelligence
Renmin University of China
Beijing, China
jrwen@ruc.edu.cn

ABSTRACT

Unbiased learning to rank methods have been proposed to address biases in search ranking. These biases, known as position bias and sample selection bias, often occur simultaneously in real applications. Existing approaches either tackle these biases separately or treat them as identical, leading to incomplete elimination of both biases. This paper employs a causal graph approach to investigate the mechanisms and interplay between position bias and sample selection bias. The analysis reveals that position bias is a common confounder bias, while sample selection bias falls under the category of collider bias. These biases collectively introduce a cascading process that leads to biased clicks. Based on our analysis, we propose Causal Likelihood Decomposition (CLD), a unified method that effectively mitigates both biases in top- k learning to rank. CLD removes position bias by leveraging propensity scores and then decomposes the likelihood of selection biased data into sample selection bias term and relevance term. By maximizing the overall log-likelihood function, we obtain an unbiased ranking model from the relevance term. We also extend CLD to pairwise neural ranking. Extensive experiments demonstrate that CLD and its pairwise neural extension outperform baseline methods by effectively mitigating both position bias and sample selection bias. The robustness of CLD is further validated through empirical studies considering variations in bias severity and click noise.

*Jun Xu is the corresponding author. Work partially done at Engineering Research Center of Next-Generation Intelligent Search and Recommendation, Ministry of Education.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR-AP '23, November 26–28, 2023, Beijing, China

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0408-6/23/11...\$15.00
<https://doi.org/10.1145/3624918.3625340>

CCS CONCEPTS

• **Information systems** → **Learning to rank**; • **Computing methodologies** → *Learning from implicit feedback*.

KEYWORDS

unbiased learning to rank, position bias, sample selection bias

ACM Reference Format:

Haiyuan Zhao, Jun Xu, Xiao Zhang, Guohao Cai, Zhenhua Dong, and Ji-Rong Wen. 2023. Unbiased Top- k Learning to Rank with Causal Likelihood Decomposition. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (SIGIR-AP '23)*, November 26–28, 2023, Beijing, China. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3624918.3625340>

1 INTRODUCTION

In order to efficiently make use of user interaction data in learning of ranking models, studies on alleviating biases in user interaction data have been conducted, called Unbiased Learning to Rank (ULTR) [4, 5, 18] or Counterfactual Learning to Rank (CLTR) [1, 19, 27]. Previously, studies focused on position bias and usually assumed that users can examine the whole ranking list so that every relevant document is guaranteed to be examined [7, 12, 23, 47]. Due to the limitation of the device sizes, however, search engines usually only display at most k relevant documents to the user-issued query, on the basis of existing ranking models. It leads to the problem of unbiased top- k learning to rank [24]. If the ranking models are trained on the user interactions with these top- k displayed documents, the sample selection bias will occur [16, 17]. Moreover, the user interaction data gathered from the top- k ranking positions still suffers from the effect of position bias, making unbiased top- k learning to rank more challenging.

Recently, Oosterhuis and de Rijke [25] developed policy-aware propensity scoring to eliminate sample selection bias. They proved that the policy-aware estimator is unbiased if every relevant item has a non-zero probability to appear in the top- k ranking. However, they treated position bias and sample selection bias as identical biases and applied a uniform policy-aware propensity score to

reweight them. Moreover, their approach relies on knowledge of multiple logging policies and requires these policies to produce sufficiently different ranking orders, which is an expensive external intervention. In a similar vein, inspired by Heckman’s two-stage method [16, 17], Ovaisi et al. [28] proposed Heckman^{rank} for top- k learning to rank. To jointly correct both sample selection bias and position bias, they introduced RankAgg, which combines the results from Heckman^{rank} and IPW (Inverse Propensity Weighting). However, as position bias and sample selection bias were mitigated separately without considering their associations, the approach may still produce biased ranking outcomes. Therefore, effectively mitigating both position bias and sample selection bias simultaneously remains a challenging problem.

In this study, we analyze the biases present in user interactions using a causal graph framework. By decomposing the likelihood of top- k ranking, we demonstrate that position bias is generated by the influence of examining items in different positions and represents a typical confounder bias. Conversely, sample selection bias can be understood as a collider bias. Furthermore, we identify an association between position bias and sample selection bias in the context of top- k ranking. The probability of a click on a document is influenced by both its displayed position and the search engine’s selection process. These findings highlight that addressing each bias separately or treating them as identical cannot effectively produce unbiased ranking outcomes. Consequently, we are motivated to develop a novel approach capable of simultaneously mitigating both biases.

Based on our analysis, we propose a unified approach, called Causal Likelihood Decomposition (CLD), to simultaneously mitigate position bias and sample selection bias. CLD follows a cascade process to address both biases effectively. First, CLD directly applies examination propensity scores to reweight observed user interactions, reducing the impact of position bias. However, these reweighted interactions still suffer from sample selection bias. To further tackle sample selection bias, CLD decomposes the log-likelihood function into distinct components. One component represents an unbiased term solely based on user-perceived relevance, while the other components capture sample selection bias. This decomposition allows for detachment of relevance signals from the observed user interaction data. Theoretical analysis demonstrates that by maximizing the entire log-likelihood function, an unbiased relevance ranking model can be obtained from the unbiased term. Furthermore, we present an extension of CLD that incorporates neural networks as ranking and selection models. The parameters of these models are learned using pairwise losses, enhancing their effectiveness in dealing with bias mitigation.

CLD offers several advantages: theoretical soundness, elegant extension to pairwise neural ranking, and high accuracy in unbiased top- k learning to rank. The major contributions of this work are:

- (1) A theoretical analysis towards position bias and sample selection bias from the viewpoint of statistical causal inference;
- (2) A unified and theoretical sound approach to mitigating both position bias and sample selection bias in top- k ranking. The method is derived under likelihood maximization and can be applied to both pointwise and pairwise training;
- (3) Extensive experiments on two publicly available datasets demonstrated the effectiveness of the proposed approaches over baselines for the task of unbiased top- k learning to rank. The empirical analysis also showed the robustness of the approaches in terms of the variation of bias severity and the click noise.

2 RELATED WORK

2.1 Unbiased learning to rank

Recently, there has been a trend towards utilizing the user interaction data (e.g., the click log) as the substitute for the expert annotated relevance labels to train the ranking models in web search [18, 39, 44]. In contrast to the expert annotated labels, the user interaction data is massive, cheap, and most importantly, user-centric [1]. However, the behaviors of the users could probably be affected by some unexpected factors [21], including the display ranking position [1, 3, 4, 13, 14, 22, 23], the search engine’s selection [25, 28, 29], and others [2, 13, 35–37, 41, 43]. These factors, along with users’ true perceived relevance, impact the observational user interaction data gathered from search engines. Although relevance signal is contained inside, the interaction data cannot be directly used to train the ranking models unless those aforementioned factors are eliminated. Otherwise, a biased ranking model would be learned and hurt the user experience [10].

To mitigate biases, unbiased learning to rank has attracted a lot of research efforts recently. Most approaches focus on addressing a single bias. For example, inverse propensity weighting (IPW) has been widely discussed in many studies [1, 23] for addressing the position bias. It estimates the causal effect of examination and extracts them from the click signal directly. The top- k cut-off of search engines leads to sample selection bias. Inspired by the famous Heckman’s two-stage method [16, 17], Ovaisi et al. [28] proposed Heckman^{rank} for top- k learning to rank. Furthermore, Oosterhuis and de Rijke [25] developed policy-aware propensity scoring to eliminate sample selection bias, by assuming that policy-aware estimator knows multiple differ enough logging policies.

In the real world, various biases could occur simultaneously [10]. To address the trust bias and position bias simultaneously, Agarwal et al. [2] proposed a Bayes-IPS estimator. Affine-IPS [37] improved Bayes-IPS and achieved better performance. Ovaisi et al. [28] proposed RankAgg who ensembles the ranking results by the model for correcting position bias and the model for correcting sample selection bias. Ovaisi et al. [29] proposed PIJD which does not require the exact propensity scores and can mitigate both position bias and sample selection bias. More recently, Oosterhuis and de Rijke [26] introduced an intervention-aware estimator for integrating counterfactual and online learning to rank, which can mitigate position bias, sample selection bias, and trust bias simultaneously. Chen et al. [9] proposed AutoDebias that leverages the uniform data to learn the optimal debiasing strategy for various biases.

2.2 Causal inference in information retrieval

The generation of users’ implicit feedback in real search engines is affected by many biased factors. To make this feedback usable, causal inference has been introduced to analyze the generation procedure of users’ implicit feedback and mitigate bias inside. For

instance, Zheng et al. [46] analyzed the casual structure of popularity bias and proposed DICE to disentangle the user interest from click. Zhang et al. [45] further analyzed the causal structure of item popularity and leveraged them to enhance the performance of recommendation. Wang et al. [42] utilized the causal inference to handle the unobserved confounders in the recommendation. Wang et al. [38] proposed DecRs to dynamically regulate backdoor adjustment according to user status, thus eliminating the effect of confounders. However, few works conduct a systematical analysis for the bias in ranking from the viewpoint of causal inference.

3 ANALYZING THE BIASES IN TOP-K RANKING

3.1 Problem Formulation

The problem of unbiased top- k learning to rank can be described as follows. Given a user query q and K retrieved documents, each query-document pair (q, d) is described by a feature vector $\mathbf{x} = \phi(q, d) \in \mathbb{R}^n$. The relevance of (q, d) can be represented by an unobserved variable R , which could be binary, ordinal, or real. The retrieved documents are ranked by a logging policy (an existing ranking model) $\pi_0 : \mathbb{R}^n \mapsto \{1, 2, \dots, K\}$ according to their features, where each document will be ranked at some position $P \in \{1, 2, \dots, K\}$ by π_0 . In the real world, only the top $k \leq K$ documents can be presented to users due to some limitations (e.g., screen sizes). Let's use $S \in \{0, 1\}$ to denote whether a document is selected and presented to the user, i.e., $S = 1$ if selected otherwise not. Further, let's use $E \in \{0, 1\}$ to denote that the user has examined the presented document, and $C \in \{0, 1\}$ to denote whether a user clicks the document, which is a random variable obeying Bernoulli distribution[1, 23, 37].

The user interactions with a search engine can be recorded as click log $\mathcal{D} = \{(\mathbf{x}_i, c_i, k_i, s_i)\}_{i=1}^N$, where $\mathbf{x}_i, c_i, k_i, s_i$ respectively denote the i -th query-document pair's feature vector, whether the document being clicked, the rank position, and whether being selected. Ideally, we hope an unbiased ranking model could be estimated by maximizing the log-likelihood shown below:

$$\mathcal{L}_{\text{unbiased}} = \sum_{i=1}^N \log(\Pr(r_i | \mathbf{x}_i)), \quad (1)$$

where the r_i is the unobserved relevance for query-document pairs encoded by \mathbf{x}_i . Equation (1) cannot be maximized because r_i cannot be observed directly.

On the other hand, we observed that the click log consists of two parts: $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_u$, where $\mathcal{D}_s = \{(\mathbf{x}_i, c_i, k_i, s_i = 1)\}_{i=1}^{N_s}$ are the interactions for the N_s selected (q, d) pairs, and $\mathcal{D}_u = \{(\mathbf{x}_i, c_i = 0, k_i, s_i = 0)\}_{i=1}^{N_u}$ are those for the N_u not selected pairs. A naive log-likelihood can be written as:

$$\begin{aligned} \mathcal{L}_{\text{naive}} = & \sum_{i=1}^{N_s} \log(\Pr(c_i | s_i = 1, \mathbf{x}_i) \Pr(s_i = 1 | \mathbf{x}_i)) \\ & + \sum_{i=1}^{N_u} \log(\Pr(s_i = 0 | \mathbf{x}_i)). \end{aligned} \quad (2)$$

Though it can be optimized, the naive log-likelihood suffers from both position bias from c_i and the sample selection bias from s_i . There exists a large gap between the naive Equation (2) and the ideal

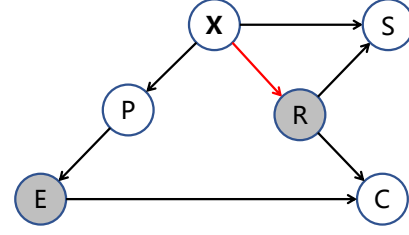


Figure 1: The casual graphs of observed log data. Each node corresponds a casual variable and the gray node means that the variable is unobserved. The red arrow ($x \rightarrow R$) denote the effect that an unbiased ranking model needs to estimate.

Table 1: Notations and explanations.

Notation	Description
(q, d)	a query-document pair
$\mathbf{x} = \phi(q, d)$	feature vector in \mathbb{R}^n corresponding to a (q, d) pair
R, r_i	true relevance of a (q, d) pair (unobserved)
E	user's examination on a document (unobserved)
C, c_i	click on a document (can be observed)
P, k_i	position of a document being displayed (observed)
S, s_i	whether a document being selected (observed)
$\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_u$	click log for selected and not selected (q, d) pairs
ρ_i	the propensity score of the i -th (q, d) pair
π_0	logging policy (an existing ranking model)

unbiased objective Equation (1). Table 1 lists the major notations in the paper.

3.2 Causal view of the biases in top- k ranking

Next, we will illustrate why ranking models are biased if they are trained with user clicks directly, and reveal the distinctions and association between the position bias and sample selection bias. In order to accomplish this, we introduce a causal graph representing the observed log data, depicted in Figure 1. This causal graph comprises six causal variables denoted as $\{x, P, R, S, E, C\}$ which have been defined in Table 1. The graph's edges describe causal relations between variables:

- $x \rightarrow R$: this edges represents the causal relations of the query- document pair's feature x and their corresponding relevance score R , which is the effect that an unbiased ranking model needs to estimate.
- $x \rightarrow P$: the displayed position P is determined by a logging policy $\pi_0(x)$ which takes the feature x as the input.
- $P \rightarrow E$: the chance of an item examined by users is determined by its displayed position P , note that the variable E is unobserved.
- $(x, R) \rightarrow S$: the rank position used to select is determined by the predicted score of logging policy π_0 , and π_0 is trained by x and R^1 , thus the selection S is indirectly determined by x

¹In real search practices, a small number of the human label can be utilized to train π_0 .

and R . The intermediate factors are omitted for simplifying the illustration.

- $(E, R) \rightarrow C$: the click C on a query-document pair will be impacted by both the examination E and the perceived relevance R of a user, which is followed by the examination hypothesis [34]

Drawing upon causal inference principles [30, 31], we identify two types of biases present in our causal graph, aligning with the previously mentioned position bias and sample selection bias:

- $E \leftarrow \mathbf{x} \rightarrow C$ (*confounder bias*): The feature vector \mathbf{x} of query-document pair is the confounder of examination E and click C , which leads to a spurious correlation between E and C . Since E is determined by rank position, the confounder bias is presented as position bias in top- k ranking.
- $\mathbf{x} \rightarrow S \leftarrow R$ (*collider bias*): S is the collider [30] between \mathbf{x} and R . The backdoor path between \mathbf{x} and R will be opened when conditioned on S . The collider bias manifests as sample selection bias in top- k ranking [11].

It is worth noting that some studies also discuss the trust bias [2, 37], which introduces an additional causal relation $P \rightarrow C$. However, trust bias can also be considered a form of confounder bias, sharing a similar mechanism with position bias. Since our primary focus is on addressing both confounder bias and collider bias, we have chosen to omit the trust bias in this study. From the analysis presented above, it becomes evident that position bias and sample selection bias in top- k ranking are distinct biases that require separate approaches for mitigation.

Meanwhile, the position bias and sample selection bias also have association according to Figure 1. Note that the existence of causal relations $R \rightarrow C$ bridge the association of the path of position bias ($E \leftarrow \mathbf{x} \rightarrow C$) and sample selection bias ($\mathbf{x} \rightarrow S \leftarrow R$). This implies that the relevance R initially experiences sample selection bias, which subsequently amplifies the spurious correlation introduced by position bias through $R \rightarrow C$. As a result, C becomes influenced by both biases simultaneously. To provide further clarity, the selected click probability $\Pr(C = 1 \mid S = 1, \mathbf{x})$ (a component of Equation (2), which is abbreviated as $p_C^{S=1}$ in the following equation) can be decomposed as follows:

$$\begin{aligned}
 p_C^{S=1} &= \sum_E \sum_R \Pr(C = 1 \mid E, R) \Pr(R \mid S = 1, \mathbf{x}) \sum_{i=1}^K \Pr(E \mid P = i) \Pr(P = i \mid \mathbf{x}) \\
 &= \sum_E \sum_R \Pr(C = 1 \mid E, R) \Pr(R \mid S = 1, \mathbf{x}) \Pr(E \mid P = k) \\
 &= \sum_R \Pr(C = 1 \mid E = 1, R) \Pr(R \mid S = 1, \mathbf{x}) \Pr(E = 1 \mid P = k) \\
 &= \underbrace{\Pr(E = 1 \mid P = k)}_{\text{Position Bias}} \underbrace{\mathbb{E}_{R \sim \Pr(R \mid S=1, \mathbf{x})} [\sigma(R)]}_{\text{Selection biased relevance}}, \quad (3)
 \end{aligned}$$

where $k = \pi_0(\mathbf{x})$. The first line is an expansion based on Figure 1. The second line is based on the assumption that π_0 is a deterministic ranking policy, meaning that one document in a query only has one corresponding rank position. The third line is according to the examination hypothesis [34]. In the last line, we treat the click probability $\Pr(C = 1 \mid E = 1, R)$ as a function of relevance R , denote

as $\sigma(R)$. Here $\sigma(\cdot)$ is a monotonically increasing function for mapping R to the interval 0 to 1 (e.g. Sigmoid function). This mapping serves as an indicator for relevance R . The decomposition equation (Equation 3) demonstrates that user-perceived relevance is initially impacted by sample selection bias and subsequently influenced by position bias, forming a cascade process. To obtain an unbiased relevance signal, it is essential to first eliminate the position bias in the click signal and then mitigate the sample selection bias.

4 UNIFIED BIAS MITIGATION IN TOP- k RANKING

Based on the analysis in the above section, this section presents a model called Causal Likelihood Decomposition (CLD) which simultaneously mitigates the position bias and the sample selection bias in top- k learning to rank.

4.1 Formulation of the log-likelihood

The analysis of Equation (3) in Section 3.2 indicates that the click signals can be transformed to selection biased relevance with the help of examination propensity $\Pr(E = 1 \mid P = k)$:

$$\mathbb{E}_{R \sim \Pr(R \mid S=1, \mathbf{x})} [\sigma(R)] = \frac{\Pr(C = 1 \mid S = 1, \mathbf{x})}{\Pr(E = 1 \mid P = k)} = \mathbb{E} \left[\frac{C}{\Pr(E = 1 \mid P = k)} \right]$$

Therefore, the click in Equation (2) can be replaced with the expectation of propensity re-weighted click, achieving a log-likelihood with position bias be detached:

$$\mathcal{L}_{\text{de. pos.}} = \sum_{i=1 \wedge s_i=1}^{N_s} \log \left(\Pr \left(\mathbb{E} \left[\frac{c_i}{\rho_i} \right], s_i \mid \mathbf{x}_i \right) \right) + \sum_{i=1 \wedge s_i=0}^{N_u} \log (\Pr(s_i \mid \mathbf{x}_i)), \quad (4)$$

where $\mathbb{E} \left[\frac{c_i}{\rho_i} \right] = \mathbb{E} \left[\frac{C}{\Pr(E=1 \mid P=k)} \right]$ for simplifying the notations. Note that ρ_i is the examination propensity score for the i -th (q, d) pair. A number of studies have been proposed to estimate the weights [3, 4, 40]. In this paper, we treat them as known values.

Equation (4) still suffers sample selection bias because the first term still contains s_i . Therefore, the likelihood of selected biased data can be further decomposed to contain the unbiased learning target, thus detaching sample selection bias:

$$\begin{aligned}
 \mathcal{L}_{\text{de. both biases}} &= \sum_{i=1 \wedge s_i=1}^{N_s} \log \left(\Pr \left(\mathbb{E} \left[\frac{c_i}{\rho_i} \right] \mid \mathbf{x}_i \right) \right) \\
 &+ \sum_{i=1 \wedge s_i=1}^{N_s} \log \left(\Pr \left(s_i \mid \mathbb{E} \left[\frac{c_i}{\rho_i} \right], \mathbf{x}_i \right) \right) \\
 &+ \sum_{i=1 \wedge s_i=0}^{N_u} \log (\Pr(s_i \mid \mathbf{x}_i)), \quad (5)
 \end{aligned}$$

Same as the naive likelihood in Equation (2), Equation (5) is still the likelihood among all observed data. The difference is both position bias and sample selection bias have been detached, and the unbiased learning target has been decomposed as the first term of Equation (5).

Algorithm 1: The training procedure of CLD

Input: iteration number T , click log $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_u$
Output: Model parameters β and ω

- 1 $\rho \leftarrow$ estimate K propensity scores;
- 2 $\beta, \omega \leftarrow$ Xavier initialization[15];
- 3 **for** $1 \leq t \leq T$ **do**
- 4 Randomly sample a batch of sessions \mathcal{D}' from $\mathcal{D}_s \cup \mathcal{D}_u$;
- 5 **for** $(\mathbf{x}_i, c_i, k_i, s_i) \in \mathcal{D}'$ **do**
- 6 $\rho_i \leftarrow \rho[k_i]$
- 7 **if** $s_i = 1$ **then**
- 8 | Update β, ω with the gradient of Eq. (6);
- 9 **else**
- 10 | Update ω with the gradient of Eq. (6);
- 11 **end**
- 12 **end**
- 13 **end**
- 14 **return** β, ω

4.2 Optimization

To optimize Equation (5), we followed the *Type II Tobit model* [6], which parameterized the likelihood in Equation (5) under the linear and Gaussian assumptions. Specifically, assuming that both the selection model and the ranking model are linear:

$$s_i = \begin{cases} 0 & \text{if } \mathbf{x}_i^T \omega + \epsilon_i \leq 0 \\ 1 & \text{if } \mathbf{x}_i^T \omega + \epsilon_i > 0 \end{cases}, \quad r_i = \begin{cases} \mathbf{x}_i^T \beta + \mu_i & \text{if } s_i = 1 \\ \text{unobserved} & \text{if } s_i = 0 \end{cases}$$

where s_i and r_i are the selection indicator and the selection biased relevance of the i -th (q, d), respectively, and both of them are calculated based on the feature vector \mathbf{x}_i . ω and β are the parameters of these two linear models. ϵ_i and μ_i are the I.I.D. noises that obey Gaussian distributions and their variances are assumed to be 1.

According to the derivations and conclusions in [6], the parameterized log-likelihood of Equation (5) becomes:

$$\begin{aligned} \mathcal{L}_{\text{CLD}}(\beta, \omega) = & - \sum_{i=1 \wedge s_i=1}^{N_s} \left(\mathbb{E} \left[\frac{c_i}{\rho_i} \right] - \mathbf{x}_i^T \beta \right)^2 \\ & + \sum_{i=1 \wedge s_i=1}^{N_s} \log \Phi \left(\frac{\mathbf{x}_i^T \omega + \gamma \left(\mathbb{E} \left[\frac{c_i}{\rho_i} \right] - \mathbf{x}_i^T \beta \right)}{(1 - \gamma^2)^{\frac{1}{2}}} \right) \\ & + \sum_{i=1 \wedge s_i=0}^{N_u} \log \left(1 - \Phi(\mathbf{x}_i^T \omega) \right), \end{aligned} \quad (6)$$

where Φ is the cumulative distribution function of a standard normal distribution, γ is the correlation coefficient of the error terms of ϵ_i and μ_i , which indicates how the selection of a (q, d) pair related to its relevance. In our implementation, γ is treated as a hyper parameter. Maximizing Equation (6) achieves an unbiased estimation of β and ω :

$$(\beta^*, \omega^*) \leftarrow \arg \max_{\beta, \omega} \mathcal{L}_{\text{CLD}}(\beta, \omega).$$

Algorithm 1 shows the procedure of the CLD learning algorithm for learning unbiased relevance ranking model β (and the selection model ω). The inputs to the algorithm are click log with feature,

Algorithm 2: Pairwise training for CLD

Input: iteration number T , click log $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_u$
Output: Model parameters β, ω

// Create preference pairs based on \mathcal{D} ;

- 1 $\rho \leftarrow$ estimate K propensity scores;
- 2 $\mathcal{D}_s^{\text{pair}} \leftarrow \left\{ ((\mathbf{x}_i, s_i), (\mathbf{x}_j, s_j)) \mid \mathbb{E} \left[\frac{c_i}{\rho[k_i]} \right] > \mathbb{E} \left[\frac{c_j}{\rho[k_j]} \right], s_i = 1 \wedge s_j = 1 \right\}$;
- 3 $\mathcal{D}_u^{\text{pair}} \leftarrow \left\{ ((\mathbf{x}_i, s_i), (\mathbf{x}_j, s_j)) \mid s_i = 0 \vee s_j = 0 \right\}$;
- 4 $\beta, \omega \leftarrow$ Xavier initialization[15];
- 5 **for** $1 \leq t \leq T$ **do**
- 6 Randomly sample a batch \mathcal{D}' from $\mathcal{D}_s^{\text{pair}} \cup \mathcal{D}_u^{\text{pair}}$;
- 7 **for** $((\mathbf{x}_i, s_i), (\mathbf{x}_j, s_j)) \in \mathcal{D}'$ **do**
- 8 **if** $s_i = 1 \vee s_j = 1$ **then**
- 9 | Update β, ω with the gradient of Eq. (9);
- 10 **else**
- 11 | Update ω with the gradient of Eq. (9);
- 12 **end**
- 13 **end**
- 14 **end**
- 15 **return** β, ω

selection indicator, and propensity score re-weighted click signals. After sampling a batch of data, the algorithm updates both models if this data record was selected, and only updates the selection model if it was not selected.

4.3 Online ranking

The outputs of the learning algorithm are the parameters of the ranking model β^* and parameters of the selection model ω^* . Intuitively, the selection model is used to absorb the sample selection bias while the relevance model is used to obtain the unbiased estimation of relevance. Therefore, in online ranking, a (q, d) pair's ranking score is calculated as an unbiased estimation of relevance:

$$\hat{r} = \langle \phi(q, d), \beta^* \rangle.$$

5 EXTENSION TO PAIRWISE NEURAL RANKING

The models learned by Algorithm 1 are limited to be linear and learned with a pointwise objective function. Previous studies have shown that the neural ranking models learned with a pairwise objective such as BPR [33] usually achieve better results. In this section, we extend the proposed pointwise and linear CLD model to pairwise neural ranking, denoted as CLD^{pair}.

To derive the pairwise format of CLD, we first give the unbiased log-likelihood in pairwise format:

$$\mathcal{L}_{\text{unbiased}}^{\text{pair}} = \sum_{r_i > r_j} \log \left(\Pr(r_i > r_j | \mathbf{x}_i, \mathbf{x}_j) \right). \quad (7)$$

For a document i and document j in the ranking list of query q , the pairwise unbiased likelihood is consists of the relative order of their relevance comparisons. Unfortunately, the relevance r_i and r_j is unknown for us. What we can observe is the click signal of each document, Based on Equation (5) and Equation (7), the decomposed

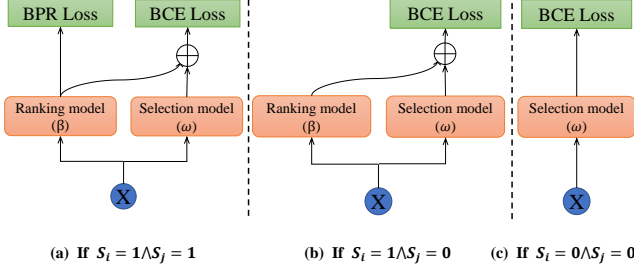


Figure 2: The unified debiasing model structure for optimizing pairwise neural format of CLD.

log-likelihood in pairwise can be written as

$$\begin{aligned} \mathcal{L}_{\text{de. both biases}}^{\text{pair}} &= \sum_{\bar{r}_i > \bar{r}_j, s_i = 1 \wedge s_j = 1} \log (\Pr(\bar{r}_i > \bar{r}_j | \mathbf{x}_i, \mathbf{x}_j)) \\ &+ \sum_{\bar{r}_i > \bar{r}_j, s_i = 1 \wedge s_j = 0} \log (\Pr(s_i, s_j | \bar{r}_i > \bar{r}_j, \mathbf{x}_i, \mathbf{x}_j)) \\ &+ \sum_{s_i = 0 \vee s_j = 0} \log (\Pr(s_i, s_j | \mathbf{x}_i, \mathbf{x}_j)), \end{aligned} \quad (8)$$

where $\bar{r}_i = \mathbb{E}[c_i / \rho_i]$ and $\bar{r}_j = \mathbb{E}[c_j / \rho_j]$. Since $\mathbb{E}[c_i / \rho_i] = r_i$, the first term of Equation (8) implies an unbiased log-likelihood. Maximizing Equation (8) can obtain an unbiased ranking model.

To conduct the optimization, we first parameterize the models with neural networks, as shown in Figure 2. Given a (q, d) pair, its representation is denoted as \mathbf{x} . Based on the representation, the relevance ranking model and selection model are defined as feed-forward neural networks, denoted as $f_{\beta}(\cdot)$ and $f_{\omega}(\cdot)$, respectively. Furthermore, we assume that in the second and third terms of Equation (8), the selection of document pairs $(\mathbf{x}_i, \mathbf{x}_j)$ are independent:

$$\begin{aligned} \Pr(s_i, s_j | \bar{r}_i > \bar{r}_j, \mathbf{x}_i, \mathbf{x}_j) &= \Pr(s_i | \bar{r}_i > \bar{r}_j, \mathbf{x}_i) \Pr(s_j | \bar{r}_i > \bar{r}_j, \mathbf{x}_j); \\ \Pr(s_i, s_j | \mathbf{x}_i, \mathbf{x}_j) &= \Pr(s_i | \mathbf{x}_i) \Pr(s_j | \mathbf{x}_j). \end{aligned}$$

As shown in Figure 2(a), if both of the documents in a pair are selected into the top- k positions, the likelihood of relevance part and conditional selection part can be formulated with BPR loss and Binary Cross Entropy loss, respectively. if only one document in a pair is selected, the conditional selection likelihood can be approximated as that shown in Figure 2(b). If neither of the two documents in a pair is selected, the selection likelihood can be formulated with Binary Cross Entropy loss directly (Figure 2(c)). Therefore, the overall pairwise objective function becomes:

$$\begin{aligned} O_{\text{CLD}}^{\text{pair}}(\boldsymbol{\beta}, \boldsymbol{\omega}) &= s_i s_j \log \sigma \left(f_{\beta}(\mathbf{x}_i) - f_{\beta}(\mathbf{x}_j) \right) \\ &+ s_i \log \sigma \left(f_{\omega}(\mathbf{x}_i) + f_{\beta}(\mathbf{x}_i) - f_{\beta}(\mathbf{x}_j) \right) + (1 - s_i) \log \sigma \left(1 - f_{\omega}(\mathbf{x}_i) \right) \\ &+ s_j \log \sigma \left(f_{\omega}(\mathbf{x}_j) + f_{\beta}(\mathbf{x}_i) - f_{\beta}(\mathbf{x}_j) \right) + (1 - s_j) \log \sigma \left(1 - f_{\omega}(\mathbf{x}_j) \right), \end{aligned} \quad (9)$$

where σ denotes the sigmoid function. To maximize Equation (9), we can get the approximately unbiased estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$:

$$(\boldsymbol{\beta}^*, \boldsymbol{\omega}^*) \leftarrow \arg \max_{\boldsymbol{\beta}, \boldsymbol{\omega}} O_{\text{CLD}}^{\text{pair}}(\boldsymbol{\beta}, \boldsymbol{\omega}).$$

Algorithm 2 illustrates the optimization procedure for Equation (9).

As for online ranking, given a (q, d) pair, its ranking score is calculated by the ranking model f_{β^*} :

$$\hat{r} = f_{\beta^*}(\phi(q, d)).$$

6 EXPERIMENT SETUP

We conducted experiments to evaluate the proposed CLD and its extension CLD^{pair} , by following the settings presented in the existing unbiased learning to rank studies [1, 4, 23, 25].

Datasets: Two widely used public datasets, YahooC14B [8] and WEB10K [32], were used in our experiments. YahooC14B contains around 30,000 queries, each associated with averaged of 24 documents. Each query-document pair is depicted with a 700-dimension feature vector and five-grade relevance labels. WEB10K has 10,000 queries and each associated with about 125 documents. Each query-document pair is depicted with a 136-dimension feature vector and a five-grade relevance label. Following the practices in [23], we converted the relevance label in both two datasets with $r = 1$ for grades 3 and 4 and $r = 0$ for the others. Only the set 1 of YahooC14B and the first fold of WEB10K was used for training. Expert annotated labels in the test sets were used to evaluate the ranking accuracy.

Click simulation: Following the practices in [23], the users' interactions with search engines were simulated and got the clicks. First, 1% labeled data were randomly sampled from the dataset and used to train an SVM^{rank} [20] as the production ranker. Then for each click session, a query was uniformly sampled and the ranking result was generated by the production ranker. To simulate users' click, the position based model (PBM) [5, 34] was adopted in which $(E = 1 \wedge R = 1) \Rightarrow C = 1$, a click occurs only when the document is examined and is relevant. For every (q, d) pair, the examination probability is based on the displayed position:

$$\Pr(E = 1 | P = k) = \begin{cases} \left(\frac{1}{k}\right)^{\eta}, & \text{if } k \leq K \\ 0, & \text{else} \end{cases} \quad (10)$$

where η is the parameter to control the severity of position bias, and K is the cut-off position. The examination probability is also the propensity score in the proposed approach and we assume it is known in advance. During the process, the irrelevant documents were allowed to be clicked with a small probability to simulate the click noise.

Baselines: State-of-the-art unbiased learning to rank approaches were adopted as the baselines:

Naive : Directly regarding the clicks as relevance labels.

IPS [23] : Correcting the position bias with propensity score.

Heckman^{rank} [28] : Correcting the sample selection bias with Heckman two-stage method.

RankAgg[28] : Mitigating both the position bias and sample selection bias by combining the results of IPS and Heckman^{rank}.

Oracle : Using the non-discarded expert annotated labels to learn the ranking model. It showed the (theoretical) performance upper bound on the dataset.

Table 2: Ranking accuracy on YaHooC14B and WEB10K. Bold-face means the best performed approaches (excluding Oracle). Experimental settings: top-5 cut-off, $\eta = 0.1$, 10^5 click sessions, and 10% click noise. We also present the 90% confidence interval of t -distribution for our methods.

Method	YaHooC14B			WEB10K		
	NDCG@1	NDCG@3	MAP	NDCG@1	NDCG@3	MAP
Naive	0.606	0.593	0.592	0.383	0.350	0.294
IPS	0.650	0.619	0.609	0.413	0.368	0.282
Heckman ^{rank}	0.608	0.590	0.587	0.350	0.331	0.287
RankAgg	0.649	0.623	0.608	0.413	0.375	0.299
CLD	0.661 ± .002	0.631 ± .001	0.616 ± .001	0.434 ± .005	0.391 ± .003	0.312 ± .001
CLD ^{pair}	0.662 ± .001	0.630 ± .001	0.615 ± .001	0.439 ± .001	0.397 ± .001	0.312 ± .000
Oracle	0.666	0.636	0.622	0.455	0.416	0.332

Policy-aware IPS [25] was not chosen as a baseline because it assumes the previous ranking models should be stochastic, which violates the Assumption (1).

Evaluation metric: NDCG@1, NDCG@3, and MAP were used to evaluate the accuracy of the baselines and the proposed method.

Implementation details: Similar to existing studies [1, 4, 37], we used a three layers neural networks with *elu* activation function as the ranking model for Naive, IPS, Oracle and CLD^{pair}, with the hidden sizes [256, 128, 64], and dropout probability of 0.5. For Heckman^{rank} and CLD (pointwise and linear), the ranking model was set to linear. The selection models for CLD and CLD^{pair} were also set to linear. The learning rate were tuned among $\{2e-4, 5e-4, 1e-3, 2e-3, 5e-3\}$. The L_2 regularization was used and the trade-off factor was tuned between $[1e-3, 1e-2]$. The correlation γ in Equation (6) was tuned between $[0.05, 0.30]$. In all of the experiments, the reported numbers were the averaged results after training 12 epochs with 5 different random seeds.

The source code, data, and experiments will be available at <https://github.com/hyz20/CLD.git>

7 RESULTS AND DISCUSSIONS

Table 2 shows the ranking accuracy of our approaches and the baselines, on YaHooC14B and WEB10K. The results showed that the proposed CLD and CLD^{pair} outperformed the baselines in terms of NDCG and MAP. ‘Oracle’ is the upper bound of the performance, since it uses expert annotated labels. The results verified the effectiveness of the unified bias mitigation in top- k ranking.

To further reveal how CLD and CLD^{pair} outperformed the baselines, we conducted a group of exploratory experiments to answer the following research questions:

- RQ1** How does CLD perform under different severity levels of sample selection bias and position bias?
- RQ2** How does CLD perform under different scales of click data?
- RQ3** Is CLD robust to click noise?
- RQ4** Is CLD robust to misspecified propensity score?
- RQ5** How does CLD perform with different base model?

7.1 The effect of biases severity (RQ1)

To varying the severity levels of sample selection bias, we changed the ranking cut-off position k from 2 to 20. Smaller k leads to more severe sample selection bias. The left two sub-figures of Figure (3) show the performance curves of different approaches w.r.t. different

k values. From the results, we can see that in general CLD and CLD^{pair} outperformed the baselines at all of the k values (except CLD when $k > 10$ on WEB10K). On both datasets, when k was small, CLD and CLD^{pair} outperformed the baselines with a large margin and achieved the performance closing to the upper bound. With the increasing of k , the improvements of CLD and CLD^{pair} over IPS gradually become limited. This is because sample selection bias gets milder for larger k , making position bias dominates the negative effects of bias. Similar performance curves also came to RankAgg, another model which can mitigate both position bias and sample selection bias.

On the contrary, increasing k will lead to the performance drop of naive methods. This is because the naive method can handle neither of these two biases. Increasing data will not further improve its performance but decrease its performance instead. Also note that CLD^{pair} outperformed CLD on WEB10K since it learns a neural ranking model based on the pairwise loss. However, all methods can achieve relatively high performances on YaHooC14B. The spaces for further improvement are limited, leading to similar performances for CLD^{pair} and CLD.

To change the severity levels of position bias, we tuned the parameter η in Equation (10) from 0.0 to 2.0. Larger η leads to more severe position bias. The right two sub-figures in Figure 3 illustrate the performances curves w.r.t. different η values. From the results, we can see that CLD and CLD^{pair} still outperformed all the baselines on both datasets. With the increasing of η , the methods that can correct position bias (except Heckman^{rank} and Naive) have slight performance drops. Among the baselines, Heckman^{rank} achieved the higher performance when $\eta = 0$ (no position bias), but dropped rapidly when η increases. RankAgg also suffered from the performance drop with the increasing η because it is an ensemble of Heckman^{rank}. The phenomenon confirmed the conclusion in Section 3.2: only mitigating one bias separately still leads to a biased result in top- k ranking. Simply aggregating the results outputted by the methods that only correct one bias still leads to biased results.

7.2 The effects of click scales (RQ2)

We tested the performances of different methods by varying the scale of the click data. Figure 4 illustrates the performance curves of different methods w.r.t. the number of click sessions used for training the models. The results indicate that both CLD and CLD^{pair} consistently outperformed the baseline methods over different click session scales. According to Equation (6) and (9), CLD and CLD^{pair} have the ability of utilizing the unobserved data in training. The ability makes them perform well even being trained with a limited number of click sessions. With more click sessions being involved in training, the performances of CLD and CLD^{pair} steadily improved. In contrast, IPS and Heckman^{rank} can only correct one bias in top- k ranking. Therefore, with the increasing number of click sessions used for training, they underperformed those methods that can mitigate both position bias and sample selection bias (e.g., RankAgg, CLD and CLD^{pair}). All these results clearly verified the advantages of the proposed unified bias mitigation.

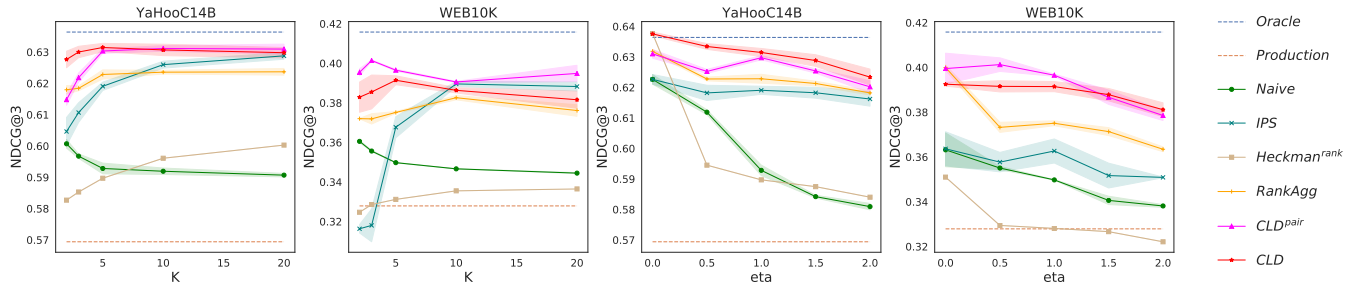


Figure 3: Performance curves of different methods w.r.t. bias severity levels. Experimental settings: 10^5 click sessions, and 10% click noise. Shaded area indicates the 90% confidence intervals of t -distribution. *Left two figures: performance curves w.r.t. different severity of sampling selection bias. Right two figures: performance curves w.r.t. different severity of position bias.*

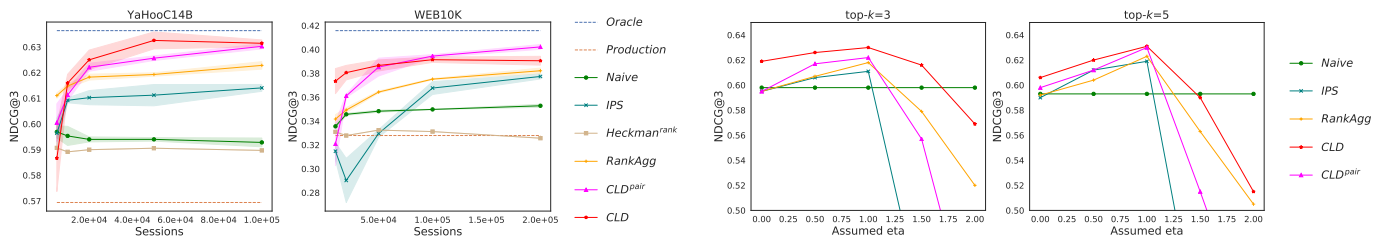


Figure 4: Performance curves of different methods w.r.t. the number of click sessions. Experimental settings: top-5 cut-off, $\eta = 1.0$ and 10% click noise.

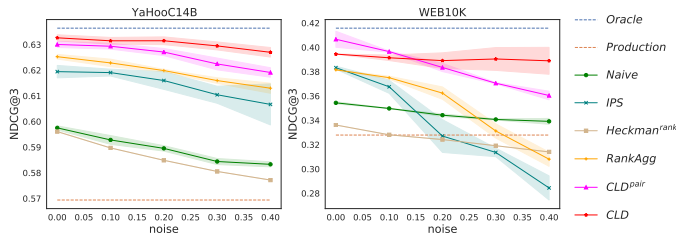


Figure 5: Performance curves of different methods w.r.t. click noise severity levels. Experimental settings: top-5 cut-off, $\eta = 1.0$ and trained with 10^5 click sessions.

7.3 The effects of click noises (RQ3)

We also conducted experiments with variant noise levels in the clicks, by changing the probability of clicking irrelevant documents from 0.0 to 0.5 when generating the training data. According to the results shown in Figure 5, both CLD and CLD^{pair} outperformed the baselines at different noise levels, indicating the robustness of the unified bias mitigation approach. Particularly, among all of the methods, CLD has minimal performance drops.

We analyzed the reasons and found that each noise click will produce more mistake pairs in pairwise methods than that of in pointwise methods. Therefore, when training with these mistake pairs, pairwise method will be suffered more. However, for pointwise method, each noise click is only presented once in the training set, making it more robust than the pairwise models.

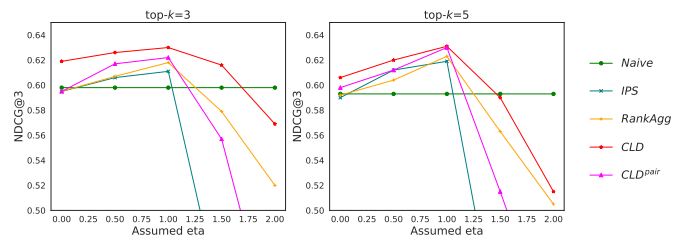


Figure 6: Performance curves of different methods on YaHooC14B w.r.t. degrees of misspecified propensity scores in different top-k cut-offs. The true $\eta = 1$ and click noise is 10%.

7.4 Effects of misspecified propensity score (RQ4)

The result we reported before assumes that the model knows the true propensity score, which is often difficult in the real world. In this experiment, we conducted experiments to test the performance of each method under various degrees on misspecified propensity scores and different top- k cut-offs, characterized by parameters η and k , respectively. The true value $\eta = 1$ and we varied it in $[0.0, 2.0]$. We tested the cases when $k = 3$ and $k = 5$. Note that Heckman^{rank} are not considered as a baseline in this experiment. This is because Heckman^{rank} does not use propensity scores.

Figure 6 illustrates the performance curves of CLD, CLD^{pair} , IPS, and RankAgg on YaHooC14B, under various degrees of misspecified propensity scores. The left and right figures respectively illustrate the results when $k = 3$ and $k = 5$. From the results, we can see that in general CLD outperformed the best in all degrees of misspecified propensity scores, which indicates the robustness of CLD. When the propensity was overestimated (i.e., $\eta < 1$), all methods related to propensity score only have a slight performance drop. However, all methods have violent performance drops if the propensity was underestimated (i.e., $\eta > 1$). This is because when the propensity is underestimated, the estimated propensity becomes smaller than its true value, and thus increasing the variance of propensity re-weighting. Even when the propensity was underestimated, the proposed CLD still outperformed other methods with large margins. This is attributed to CLD avoids dividing by propensity score in the whole loss function and therefore can reduce the variance caused by the underestimation of the propensity. Moreover, we

Table 3: Ranking accuracy comparison among different variants of CLD on YaHooC14B and WEB10K. Boldface means the best performed approaches (excluding Oracle). Experimental settings: top-5 cut-off, $\eta = 0.1$, 10^5 click sessions, and 10% click noise.

Method	YaHooC14B			WEB10K		
	NDCG@1	NDCG@3	MAP	NDCG@1	NDCG@3	MAP
CLD	0.661	0.631	0.616	0.434	0.391	0.312
CLD-N	0.652	0.619	0.610	0.340	0.321	0.284
CLD ^{pair}	0.662	0.630	0.615	0.439	0.397	0.312
CLD ^{pair} -L	0.660	0.634	0.618	0.431	0.389	0.309
Oracle	0.666	0.636	0.622	0.455	0.416	0.332

found that the effects of misspecified propensity scores were more severe on large k . This is because the larger the ranking positions, the more suffers come from the position bias.

7.5 The effects of base model (RQ5)

In previous experiments, CLD was designed as a linear ranking model because of its theoretic guarantees, while CLD^{pair} was designed to use nonlinear neural networks as its ranker. In this experiment, we modified these models so that CLD was based on a neural network with three hidden layers and CLD^{pair} used a linear model as the ranker, denoted as CLD-N and CLD^{pair}-L, respectively.

Table 3 reports the ranking accuracy of CLD, CLD^{pair}, and their variations, on YaHooC14B and WEB10K. From the results, we found that (1) CLD-N performed worst among these methods, especially on WEB10K. Compared to CLD, CLD-N used a nonlinear neural network as its ranker, which makes it lose the theoretical guarantees; (2) CLD^{pair} outperformed CLD^{pair}-L on WEB10K and performed comparably on YaHooC14B. Please note that WEB10K is larger than YaHooC14B and all methods can achieve relatively scores on YaHooC14B. We concluded that using a linear model in CLD and using nonlinear neural networks in CLD^{pair} are reasonable settings.

8 CONCLUSIONS

In this paper, we have proposed a novel and theoretical sound model for learning unbiased ranking models in top- k learning to rank, referred to as CLD. In contrast to existing methods, CLD simultaneously tackles the position bias and sampling selection biases from the viewpoint of a causal graph. It decomposes the log-likelihood function of user interactions as an unbiased relevance term plus other terms that model the biases. An unbiased ranking model can be obtained by maximizing the whole log-likelihood. Extension to the pairwise neural ranking is also developed. Experimental results verified the superiority of the proposed methods over the baselines in terms of ranking accuracy and robustness.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (No. 62376275), Beijing Outstanding Young Scientist Program NO. BJJWZYJH012019100020098, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (23XNKJ13). This work was also supported by the Intelligent Social Governance Interdisciplinary

Platform, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, Renmin University of China. The work was partially done at Beijing Key Laboratory of Big Data Management and Analysis Methods.

REFERENCES

- [1] Aman Agarwal, Kenta Takatsu, Ivan Zaitsev, and Thorsten Joachims. 2019. A General Framework for Counterfactual Learning-to-Rank. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) (SIGIR '19). ACM, New York, NY, USA, 5–14. <https://doi.org/10.1145/3331184.3331202>
- [2] Aman Agarwal, Xuanhui Wang, Cheng Li, Michael Bendersky, and Marc Najork. 2019. Addressing Trust Bias for Unbiased Learning-to-Rank. In *The World Wide Web Conference* (San Francisco, CA, USA) (WWW '19). ACM, New York, NY, USA, 4–14. <https://doi.org/10.1145/3308558.3313697>
- [3] Aman Agarwal, Ivan Zaitsev, Xuanhui Wang, Cheng Li, Marc Najork, and Thorsten Joachims. 2019. Estimating Position Bias without Intrusive Interventions. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (Melbourne VIC, Australia) (WSDM '19). ACM, New York, NY, USA, 474–482. <https://doi.org/10.1145/3289600.3291017>
- [4] Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W. Bruce Croft. 2018. Unbiased Learning to Rank with Unbiased Propensity Estimation. In *The 41st International ACM SIGIR Conference on Research Development in Information Retrieval* (Ann Arbor, MI, USA) (SIGIR '18). ACM, New York, NY, USA, 385–394. <https://doi.org/10.1145/3209978.3209986>
- [5] Qingyao Ai, Tao Yang, Huazheng Wang, and Jiaxin Mao. 2021. Unbiased Learning to Rank: Online or Offline? *ACM Trans. Inf. Syst.* 39, 2, Article 21 (Feb. 2021), 29 pages. <https://doi.org/10.1145/3439861>
- [6] T. Amemiya. 1984. Tobit models: A survey. *Journal of Econometrics* 24, 1-2 (1984).
- [7] Ben Carterette and Praveen Chandar. 2018. Offline Comparative Evaluation with Incremental, Minimally-Invasive Online Feedback. In *The 41st International ACM SIGIR Conference on Research Development in Information Retrieval* (Ann Arbor, MI, USA) (SIGIR '18). ACM, New York, NY, USA, 705–714. <https://doi.org/10.1145/3209978.3210050>
- [8] Olivier Chapelle and Yi Chang. 2011. Yahoo! learning to rank challenge overview. In *Proceedings of the learning to rank challenge*. PMLR, 1–24.
- [9] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. 2021. AutoDebias: Learning to Debias for Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) (SIGIR '21). ACM, New York, NY, USA, 21–30. <https://doi.org/10.1145/3404835.3462919>
- [10] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2020. Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240* (2020).
- [11] Stephen R Cole, Robert W Platt, Enrique F Schisterman, Haitao Chu, Daniel Westreich, David Richardson, and Charles Poole. 2010. Illustrating bias due to conditioning on a collider. *International journal of epidemiology* 39, 2 (2010), 417–420.
- [12] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An Experimental Comparison of Click Position-Bias Models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (Palo Alto, California, USA) (WSDM '08). ACM, New York, NY, USA, 87–94. <https://doi.org/10.1145/1341531.1341545>
- [13] Zhenhua Dong, Hong Zhu, Pengxiang Cheng, Xinhua Feng, Guohao Cai, Xiquiang He, Jun Xu, and Jirong Wen. 2020. Counterfactual learning for recommender system. In *Fourteenth ACM Conference on Recommender Systems*. 568–569.
- [14] Zhichong Fang, Aman Agarwal, and Thorsten Joachims. 2019. Intervention Harvesting for Context-Dependent Examination-Bias Estimation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) (SIGIR'19). ACM, New York, NY, USA, 825–834. <https://doi.org/10.1145/3331184.3331238>
- [15] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 249–256.
- [16] James J Heckman. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement*. NBER, 475–492.
- [17] James J Heckman. 1979. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* (1979), 153–161.
- [18] Ziniu Hu, Yang Wang, Qu Peng, and Hang Li. 2019. Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm. In *The World Wide Web Conference* (San Francisco, CA, USA) (WWW '19). ACM, New York, NY, USA, 2830–2836. <https://doi.org/10.1145/3308558.3313447>

- [19] Rolf Jagerman, Harrie Oosterhuis, and Maarten de Rijke. 2019. To Model or to Intervene: A Comparison of Counterfactual and Online Learning to Rank from User Interactions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) (SIGIR '19). ACM, New York, NY, USA, 15–24. <https://doi.org/10.1145/3331184.3331269>
- [20] Thorsten Joachims. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Edmonton, Alberta, Canada) (KDD '02). Association for Computing Machinery, New York, NY, USA, 133–142. <https://doi.org/10.1145/775047.775067>
- [21] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Salvador, Brazil) (SIGIR '05). ACM, New York, NY, USA, 154–161. <https://doi.org/10.1145/1076034.1076063>
- [22] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. 2007. Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. *ACM Trans. Inf. Syst.* 25, 2 (2007), 7–es. <https://doi.org/10.1145/1229179.1229181>
- [23] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) (WSDM '17). ACM, New York, NY, USA, 781–789. <https://doi.org/10.1145/3018661.3018699>
- [24] Shuzi Niu, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2012. Top-k Learning to Rank: Labeling, Ranking and Evaluation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Portland, Oregon, USA) (SIGIR '12). ACM, New York, NY, USA, 751–760. <https://doi.org/10.1145/2348283.2348384>
- [25] Harrie Oosterhuis and Maarten de Rijke. 2020. Policy-Aware Unbiased Learning to Rank for Top-k Rankings. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) (SIGIR '20). ACM, New York, NY, USA, 489–498. <https://doi.org/10.1145/3397271.3401102>
- [26] Harrie Oosterhuis and Maarten de Rijke. 2021. Unifying Online and Counterfactual Learning to Rank: A Novel Counterfactual Estimator That Effectively Utilizes Online Interventions. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) (WSDM '21). ACM, New York, NY, USA, 463–471. <https://doi.org/10.1145/3437963.3441794>
- [27] Harrie Oosterhuis and Maarten de Rijke. 2021. Robust Generalization and Safe Query-Specialization in Counterfactual Learning to Rank. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) (WWW '21). Association for Computing Machinery, New York, NY, USA, 158–170. <https://doi.org/10.1145/3442381.3450018>
- [28] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for Selection Bias in Learning-to-Rank Systems. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) (WWW '20). ACM, New York, NY, USA, 1863–1873. <https://doi.org/10.1145/3366423.3380255>
- [29] Zohreh Ovaisi, Kathryn Vasilaky, and Elena Zheleva. 2021. Propensity-Independent Bias Recovery in Offline Learning-to-Rank Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) (SIGIR '21). ACM, New York, NY, USA, 1763–1767. <https://doi.org/10.1145/3404835.3463097>
- [30] Judea Pearl. 2009. *Causality*. Cambridge university press.
- [31] Judea Pearl. 2016–2016. *Causal inference in statistics: a primer*. Wiley, Chichester, West Sussex.
- [32] Tao Qin and Tie-Yan Liu. 2013. Introducing LETOR 4.0 datasets. *arXiv preprint arXiv:1306.2597* (2013).
- [33] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).
- [34] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting Clicks: Estimating the Click-through Rate for New Ads (WWW '07). ACM, New York, NY, USA, 521–530. <https://doi.org/10.1145/1242572.1242643>
- [35] Yuta Saito. 2020. Asymmetric Tri-Training for Debiasing Missing-Not-At-Random Explicit Feedback. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) (SIGIR '20). ACM, New York, NY, USA, 309–318. <https://doi.org/10.1145/3397271.3401114>
- [36] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.
- [37] Ali Vardasbi, Harrie Oosterhuis, and Maarten de Rijke. 2020. When Inverse Propensity Scoring Does Not Work: Affine Corrections for Unbiased Learning to Rank. In *Proceedings of the 29th ACM International Conference on Information Knowledge Management*. ACM, New York, NY, USA, 1475–1484. <https://doi.org/10.1145/3340531.3412031>
- [38] Wenjie Wang, Fuli Feng, Xiangnan He, Xiang Wang, and Tat-Seng Chua. 2021. Deconfounded Recommendation for Alleviating Bias Amplification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining* (Virtual Event, Singapore) (KDD '21). Association for Computing Machinery, New York, NY, USA, 1717–1725. <https://doi.org/10.1145/3447548.3467249>
- [39] Xuanhui Wang, Michael Bendersky, Donald Metzler, and Marc Najork. 2016. 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* (Pisa, Italy) (SIGIR '16). ACM, New York, NY, USA, 115–124. <https://doi.org/10.1145/2911451.2911537>
- [40] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position Bias Estimation for Unbiased Learning to Rank in Personal Search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) (WSDM '18). ACM, New York, NY, USA, 610–618. <https://doi.org/10.1145/3159652.3159732>
- [41] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly robust joint learning for recommendation on data missing not at random. In *International Conference on Machine Learning*. PMLR, 6638–6647.
- [42] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. 2020. Causal Inference for Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) (RecSys '20). ACM, New York, NY, USA, 426–431. <https://doi.org/10.1145/3383313.3412225>
- [43] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining* (Virtual Event, Singapore) (KDD '21). Association for Computing Machinery, New York, NY, USA, 1791–1800. <https://doi.org/10.1145/3447548.3467289>
- [44] Bowen Yuan, Yaxu Liu, Jui-Yang Hsia, Zhenhua Dong, and Chih-Jen Lin. 2020. Unbiased Ad Click Prediction for Position-Aware Advertising Systems. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) (RecSys '20). ACM, New York, NY, USA, 368–377. <https://doi.org/10.1145/3383313.3412241>
- [45] Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) (SIGIR '21). ACM, New York, NY, USA, 11–20. <https://doi.org/10.1145/3404835.3462875>
- [46] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) (WWW '21). ACM, New York, NY, USA, 2980–2991. <https://doi.org/10.1145/3442381.3449788>
- [47] Honglei Zhuang, Zhen Qin, Xuanhui Wang, Michael Bendersky, Xinyu Qian, Po Hu, and Dan Chary Chen. 2021. Cross-Positional Attention for Debiasing Clicks. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) (WWW '21). Association for Computing Machinery, New York, NY, USA, 788–797. <https://doi.org/10.1145/3442381.3450098>