

Entire-Space Variational Information Exploitation for Post-Click Conversion Rate Prediction

Ke Fei, Xinyue Zhang, Jingjing Li*

University of Electronic Science and Technology of China, Chengdu, China
q191106702@gmail.com, qzmvcl@gmail.com, lijn117@yeah.net

Abstract

In recommender systems, post-click conversion rate (CVR) estimation is an essential task to model user preferences for items and estimate the value of recommendations. Sample selection bias (SSB) and data sparsity (DS) are two persistent challenges for post-click conversion rate (CVR) estimation. Currently, entire-space approaches that exploit unclicked samples through knowledge distillation are promising to mitigate SSB and DS simultaneously. Existing methods use non-conversion, conversion, or adaptive conversion predictors to generate pseudo labels for unclicked samples. However, they fail to consider the unbiasedness and information limitations of these pseudo labels. Motivated by such analysis, we propose an entire-space variational information exploitation framework (EVI) for CVR prediction. First, EVI uses a conditional entire-space CVR teacher to generate unbiased pseudo labels. Then, it applies variational information exploitation and logit distillation to transfer non-click space information to the target CVR estimator. We conduct extensive offline experiments on six large-scale datasets. EVI demonstrated a 2.25% average improvement compared to the state-of-the-art baselines.

Code — <https://github.com/q1179897215/EVI>

Extended version — <https://github.com/q1179897215/EVI>

1 Introduction

Modern recommender systems aim to connect users with the items they want using efficient computation. Estimating the post-click conversion rate (CVR) is a fundamental task that models the interaction between users and items, such as user preferences for specific items. (Xi et al. 2021; Ma et al. 2018b; Dai et al. 2022). The naïve CVR model is trained using clicked samples from the click space, but it needs to estimate CVR for all impression items from the entire impression space, which includes unclicked samples. Since click and conversion events are interconnected, unclicked samples are not missing at random (MNAR) (Marlin and Zemel 2009; Zhang et al. 2020). This gap between the click space and the entire space leads to sample selection bias (SSB) (Ma et al. 2018b; Li et al. 2023a). Moreover, the data sparsity (DS) of clicked samples and converted samples further

hinder the generalization of click-space CVR models to entire space (Wang et al. 2022a).

To mitigate the SSB, causal inference techniques such as inverse propensity scores (IPS) and double robust learning (DR) have been proposed to debias the CVR estimation and control variance (Zhang et al. 2020; Guo et al. 2021; Dai et al. 2022). The essence of IPS and DR is to reconstruct the entire space loss by weighting the clicked samples' losses with IPS.

Since causal methods are trained only on click space, exploring the entire space is a promising approach to mitigate SSB and DS simultaneously (Su et al. 2024). Initially, to utilize unclicked samples, Ma et al. (Ma et al. 2018b) proposed ESMM to model CVR on the entire space through learning a click-through and conversion rate (CTCVR) task. Based on the chain rule of probability, CTCVR = click-through rate (CTR) * CVR. Wang et al. (Wang et al. 2022a) combine IPS and DR with ESMM to debias the CVR estimator and simultaneously exploit the entire space information. Xu et al. (Xu et al. 2022) proposed UKD, which trains an adaptive CVR teacher through adversarial learning to generate soft labels for unclicked samples. These pseudo labels allow the CVR estimator to be trained on the entire space using a distillation approach. In DCMT (Zhu et al. 2023), Zhu et al. labels unclicked samples as non-conversions and apply inverse non-click propensity to correct the bias in the non-conversion estimators. Similar to UKD and DCMT, Su et al. (Su et al. 2024) train a naïve CVR estimator to produce pseudo labels for unclicked samples.

From a causal inference perspective (Rubin 2005), we must measure all variables that influence clicks and conversions (also called confounders) to ensure unbiased CVR estimation. However, most existing entire-space methods ignore click information (which contains confounders) when generating pseudo labels, thus transferring the bias to the target estimator. UKD employs adversarial learning to mitigate the domain shift between the click and non-click spaces, but it also results in the loss of click information. DDPO uses the naïve CVR estimator to generate pseudo CVR labels for unclicked samples, neglecting click information when training the CVR estimator.

Since the information carried by non-click space pseudo soft labels is easily biased and limited, it is beneficial to construct an unbiased CVR teacher and transfer information at

*Corresponding author
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

both the logit and feature levels. In this paper, we propose an entire-space variational information exploitation framework (EVI) for CVR prediction. Firstly, we use click propensities to condition the original sample representations. Based on the click-conditioned representations, we then train an unbiased CVR teacher. To better utilize non-click spaces, we maximize the variational information between the CVR teacher and the CVR student, and we perform logit distillation using unbiased pseudo labels. Our contributions are highlighted as follows:

- We introduce EVI, a novel entire-space approach for CVR prediction that avoids bias from pseudo labels and enhances the exploitation of non-click space using features and logits distillation.
- We validate the effectiveness of our proposed EVI on five public and one industrial dataset. In general, EVI shows an average AUC improvement of 2.25% and a 2.78% NLL reduction over the optimal baselines.
- We demonstrate both theoretically and experimentally that the unbiasedness of EVI can be achieved. Additionally, we show that conditional entire-space CVR teacher and variational information exploitation can significantly reduce the mean bias of the CVR estimate.

2 Related Works

2.1 Multi-task Learning

In recommender systems, click events directly affect conversion events, indicating the existence of common variables (confounders) that affect both. CTR and CVR tasks are jointly optimized in a unified multi-task learning (MTL) framework for exploit these confounders.

Based on the intuition from the Mixture of Experts (MoE) approach (Jacobs et al. 1991), Ma et al. (Ma et al. 2018a) proposed MMOE, which uses multiple expert networks and task-specific gates to combine task features for CTR and CVR towers. The aim of MMOE is to reduce negative transfer between tasks. While MMOE attempts to solve the issue of negative transfer, Xi et al. proposed AITM (Xi et al. 2021), which uses an attention mechanism to transfer information from clicks to conversions. Several studies (Ma et al. 2018b; Zhang et al. 2022; Wen et al. 2021, 2020) introduce special auxiliary tasks to transfer new information to the target tasks.

2.2 Causal Inference

Common multi-task frameworks exploit confounders’ information through parameter sharing or attention layers. Causal methods employ pCTR as a compression of confounders’ information to handle the SSB problem in CVR estimation (Luo et al. 2024). Causal inference techniques, such as IPS and DR (Pearl 2009) are employed to the remove the SSB (Zhang et al. 2020; Saito 2020; Guo et al. 2021; Saito et al. 2020; Wang et al. 2022b,a; Schnabel et al. 2016) in CVR task. Causal methods weight the CVR loss using predicted inverse propensity scores, which are easily affected by high variance issues. SNIP (Schnabel et al. 2016), Clipped-IPS (Saito et al. 2020), MRDR (Guo et al. 2021), DR-MSE (Dai

et al. 2022), DR-BIAS (Dai et al. 2022), MRDR-GPL (Zhou et al. 2023), and DR-V2 (Li et al. 2023b) have been proposed to control the variance.

3 Preliminary

3.1 Problem Formulation

In online recommender systems, we collect data on user attributes, item attributes, and their interactions (such as clicks and conversions) from the entire impression space, also referred to as the entire space. This data allows us to train CTR and CVR estimators simultaneously using a multi-task framework. Let $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ be the set of users and $\mathcal{I} = \{i_1, i_2, \dots, i_n\}$ be the set of items. The entire space is delineated by the set $\mathcal{D} = \mathcal{U} \times \mathcal{I}$, where \times denotes the Cartesian product operation. Let $o_{u,i}$ be the true click label for the user-item pair (u, i) , which can be either 0 or 1. Therefore, the entire space \mathcal{D} is divided into two subsets: the click space \mathcal{O} , where $o_{u,i} = 1$, and the non-click space \mathcal{N} , where $o_{u,i} = 0$. The predicted click-through rate (pCTR) or click propensity, $\hat{o}_{u,i} \in [0, 1]$, is estimated by a model $f(\cdot; \theta_{\text{CTR}})$ trained on \mathcal{D} with parameters θ_{CTR} . Let $r_{u,i} \in \{0, 1\}$ denote the true post-click labels. The predicted post-click conversion rate (pCVR), $\hat{r}_{u,i} \in [0, 1]$, is estimated by models $f(\cdot; \theta_{\text{CVR}})$ with parameters θ_{CVR} . Let $\mathbf{x}_{u,i}$ represent the feature embeddings of the user-item pair $(u, i) \in \mathcal{D}$.

3.2 Entire-Space CVR Estimation

Given that the CVR estimator is required to predict the pCVR in the entire space \mathcal{D} , the ideal loss function for the CVR estimator is:

$$\mathcal{L}_{\text{CVR}}^{\text{ideal}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(r_{u,i}, \hat{r}_{u,i}), \quad (1)$$

where $\delta(\cdot, \cdot)$ denotes a specific loss function, commonly cross-entropy loss function (Mao, Mohri, and Zhong 2023). Since the CVR labels of unclicked samples are missing, the naïve CVR estimator can only be trained using clicked samples. The training loss for the naïve CVR estimator is defined as:

$$\mathcal{L}_{\text{CVR}}^{\text{naive}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i}). \quad (2)$$

However, these CVR labels of unclicked samples are missing not at random (MNAR) (Marlin and Zemel 2009; Zhang et al. 2020), leading to SSB, i.e.,

$$|\mathbb{E}_{\mathcal{D}} [\mathcal{L}_{\text{CVR}}^{\text{naive}}] - \mathcal{L}_{\text{CVR}}^{\text{ideal}}| > 0. \quad (3)$$

SSB originates from the interdependence between clicks and conversions. This means that post-click conversion events and click events are both influenced by some common variables (confounders). From a causal inference perspective (Rubin 2005), we must measure all confounders to ensure the unbiasedness of the CVR estimator. Co-training the CTR estimator helps the CVR estimator capture confounding information, improving the debiasing effect. The

effectiveness of multi-task learning has been verified in previous works such as MMOE (Ma et al. 2018a) and AITM (Xi et al. 2021). The CTR task loss is formulated as:

$$\mathcal{L}_{\text{CTR}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(o_{u,i}, \hat{o}_{u,i}). \quad (4)$$

The ESMM (Ma et al. 2018b) addresses SSB by introducing an entire-space CTCVR task, which directly connects CVR to CTR based on the chain rule of probability (Feller 1991). The ESMM extends the naïve CVR estimator to the entire-space CVR estimator, and the CTCVR loss is formulated as:

$$\mathcal{L}_{\text{CTCVR}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(o_{u,i} * r_{u,i}, \hat{o}_{u,i} * \hat{r}_{u,i}). \quad (5)$$

ESCM² integrates IPW and DR with ESMM. The ESCM²-IPW and ESCM²-DR estimators are theoretically unbiased if the CTR estimation is accurate. The main difference between ESCM² and ESMM lies in the CVR loss:

$$\mathcal{L}_{\text{CVR}}^{\text{IPW}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i})}{\hat{o}_{u,i}}, \quad (6)$$

$$\mathcal{L}_{\text{CVR}}^{\text{DR}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{o}_{u,i}} \right], \quad (7)$$

where $e_{u,i} = \delta(r_{u,i}, \hat{r}_{u,i})$ and $\hat{e}_{u,i}$ is predicted $e_{u,i}$ by an imputation model with parameter $f(\cdot; \theta_{\text{IMP}})$.

Unlike ESMM, which uses CTCVR estimation, UKD (Xu et al. 2022) employs a click-adaptive CVR teacher to produce pseudo conversion labels $r_{u,i}^a$ for unclicked samples. UKD enables training the CVR estimator on the entire space. The CVR loss of UKD is formulated as:

$$\begin{aligned} \mathcal{L}_{\text{CVR}}^{\text{UKD}} = \\ \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} [o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i}) + (1 - o_{u,i}) \delta(r_{u,i}^a, \hat{r}_{u,i})], \end{aligned} \quad (8)$$

UKD trains a click-adaptive CVR teacher using adversarial domain adaptation techniques like gradient reversal (Ganin et al. 2016). This approach enables the CVR teacher to generate click-adaptive representations that can deceive a click discriminator, rendering it unable to distinguish between clicked and unclicked samples. However, 'click-adaptive' also implies the loss of click information, which includes all confounding information. The loss of confounding information leads to SSB in the CVR teacher, which in turn gets transferred to the CVR student (see Theorem 1).

Similar to UKD, DDPO (Su et al. 2024) constructs a naïve CVR teacher trained on click space to generate pseudo conversion labels $r_{u,i}^c$ for unclicked samples. Additionally, DDPO uses inverse click propensity scores and inverse non-click propensity scores to weight the losses of clicked and

unclicked samples, respectively.

$$\begin{aligned} \mathcal{L}_{\text{CVR}}^{\text{DDPO}} = \\ \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[\frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i})}{\hat{o}_{u,i}} + \frac{(1 - o_{u,i}) \delta(r_{u,i}^c, \hat{r}_{u,i})}{1 - \hat{o}_{u,i}} \right] \end{aligned}$$

DDPO is based on a strong assumption (see Theorem 1 in DDPO original paper) that the CVR estimate of DDPO is unbiased when pseudo labels produced CVR teacher are accurate. However, the CVR teacher of DDPO are trained on click space, which is prone to SSB.

Theorem 1. *The entire-space CVR estimator is biased when the pseudo conversion labels of unclicked samples $r_{u,i}^*$ are biased, i.e.,*

$$|\mathbb{E}_{\mathcal{D}} [\mathcal{L}_{\text{CVR}}^{\text{entire}}] - \mathcal{L}_{\text{CVR}}^{\text{ideal}}| > 0.$$

Proof. See Appendix B.1. \square

In general, the main aim of entire-space CVR modeling is to leverage the non-click space information to combat SSB and DS. However, pseudo labels carry limited information and are easily biased due to the loss of click information. In this paper, we aim to generate unbiased pseudo labels and enhance non-click space exploitation through variational information exploitation and logit distillation.

4 Proposed Method

4.1 EVI Framework

As discussed in Section 3, the pseudo labels carry limited information and are easily biased due to the loss of click information in CVR teacher. Motivated by this, we propose EVI. As shown in Figure 1, we construct an entire-space CVR teacher trained on click-conditioned representations to ensure the unbiasedness of the CVR teacher. To enhance the entire-space knowledge transfer from the CVR teacher to the target CVR estimator (student), we perform two main actions: 1) use the conditional entire-space CVR teacher to produce unbiased pseudo conversion labels for logit distillation, and 2) maximize the variational information between the CVR teacher and the CVR student.

4.2 Conditional Entire-Space CVR Teacher

In Section 3, we demonstrate that if the pseudo labels are biased, the entire-space CVR estimator will also be biased. We propose a conditional entire-space CVR teacher to generate unbiased pseudo labels. The naïve entire-space CVR teacher model CVR as $\mathbb{P}(r_{u,i} = 1)$, which is biased since users in the click space have a higher chance of conversion (Schnabel et al. 2016):

$$\mathbb{E}_{\mathcal{D}}[\mathbb{P}(r_{u,i} = 1)] \neq \mathbb{E}_{\mathcal{D}}[\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1)]. \quad (9)$$

The conditional entire-space CVR teacher model CVR as $\mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1))$. The 'do' operator, referring to do-calculus, removes the dependency from click to conversion, ensuring unbiased results (Wang et al. 2022a). In this paper, we adopt the click propensity conditioning approach, which is similar to IPS (Zhang et al. 2020). Specifically, we train a

CTR estimator to predict the click propensity, which can be used to represent all the confounders (Rubin 2005). We then project the click propensity through a linear layer and use the outer product of this projection and the CVR-T representation learner’s output to generate the click-conditioned representation, as shown in Figure 1. Based on the click-conditioned representation, we can model the teacher’s CVR as $\mathbb{P}(r_{u,i} = 1 | \hat{O}_{u,i} = \hat{o}_{u,i})$. The unbiasedness of the conditional entire-space CVR teacher and EVI CVR estimator are proven in Theorem 2 and 3.

Theorem 2. *The conditional entire-space CVR teachers are unbiased, provided that $\hat{o}_{u,i}$ is accurate, i.e.,*

$$\mathbb{E}_{\mathcal{D}}[\mathbb{P}(r_{u,i} = 1 | \hat{O}_{u,i} = \hat{o}_{u,i})] = \mathbb{E}_{\mathcal{D}}[\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1)].$$

Proof. See Appendix B.2. \square

Theorem 3. *The EVI CVR estimator is unbiased when the click propensity $\hat{o}_{u,i}$ and pseudo conversion labels of unclicked samples $r_{u,i}^*$ are accurate, i.e.,*

$$|\mathbb{E}_{\mathcal{D}}[\mathcal{L}_{\text{CVR}}^{\text{EVI}}] - \mathcal{L}_{\text{CVR}}^{\text{ideal}}| = 0.$$

Proof. See Appendix B.3. \square

4.3 Variational Information Exploitation

Existing entire-space methods use logits distillation to learn the non-click space information. However, the information carried in the logits (pseudo labels) is limited and easily biased.

From an information theory perspective, maintaining high mutual information between the teacher and student network layers can improve the transfer of non-click space information (Ahn et al. 2019). To address the computational difficulty of mutual information, we enhance non-click space exploitation based on a variational information maximization scheme (Barber and Agakov 2004; Ahn et al. 2019), which maximizes the variational lower bound instead. Passing the feature embeddings $\mathbf{x}_{u,i}$ through the networks in Figures 1 creates K pairs of random variables $\{(\mathbf{t}^{(k)}, \mathbf{s}^{(k)})\}_{k=1}^K$. These pairs represent the output of the learner layers for CVR-T (teacher) and CVR (student). The variational lower bound for the mutual information $I(\mathbf{t}; \mathbf{s})$ of each layer pair is defined as follows:

$$\begin{aligned} I(\mathbf{t}; \mathbf{s}) &= H(\mathbf{t}) - H(\mathbf{t}|\mathbf{s}) \\ &= H(\mathbf{t}) + \mathbb{E}_{\mathbf{t}, \mathbf{s}}[\log p(\mathbf{t}|\mathbf{s})] \\ &= H(\mathbf{t}) + \mathbb{E}_{\mathbf{t}, \mathbf{s}}[\log q(\mathbf{t}|\mathbf{s})] + \mathbb{E}_{\mathbf{s}}[D_{\text{KL}}(p(\mathbf{t}|\mathbf{s})||q(\mathbf{t}|\mathbf{s}))] \\ &\geq H(\mathbf{t}) + \mathbb{E}_{\mathbf{t}, \mathbf{s}}[\log q(\mathbf{t}|\mathbf{s})], \end{aligned} \tag{10}$$

where entropy $H(\mathbf{t})$ and the conditional entropy $H(\mathbf{t}|\mathbf{s})$ are calculated from the joint distribution $p(\mathbf{t}, \mathbf{s})$. D_{KL} is Kullback-Leiber divergence (Goldberger and Greenspan 2003). $q(\mathbf{t}, \mathbf{s})$ is a variational distribution that approximates $p(\mathbf{t}, \mathbf{s})$. In this work, we employ Gaussian distribution with mean $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ as the variational distribution $q(\mathbf{t}, \mathbf{s})$. We parameterize the $\boldsymbol{\mu}$ with a linear layer of parameters $\boldsymbol{\theta}_s$, i.e., $\boldsymbol{\mu} = f(\mathbf{s}; \boldsymbol{\theta}_s)$. We set $\boldsymbol{\rho}$ as trainable parameters that pass through a softplus function (Zheng et al. 2015), ensuring

their positivity. Ignoring the constant $H(\mathbf{t})$, the variational information exploitation (VIE) loss is defined as:

$$\mathcal{L}_{\text{VIE}}^{\text{EVI}} = \sum_{n=1}^N \log \sigma_n + \frac{(t_n - \mu_n(\mathbf{s}))^2}{2\sigma_n^2}, \tag{11}$$

where t_n , μ_n and ρ_n indicates the n -th entry of the vector \mathbf{t} , $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$.

4.4 Entire-space CVR Optimization

In EVI, we optimize CTR, CVR-T, CVR and CTCVR in a single model with embedding sharing. The CTR and CTCVR loss are defined as:

$$\mathcal{L}_{\text{CTR}}^{\text{EVI}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(o_{u,i}, \hat{o}_{u,i}), \tag{12}$$

$$\mathcal{L}_{\text{CTCVR}}^{\text{EVI}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(o_{u,i} * r_{u,i}, \hat{o}_{u,i} * \hat{r}_{u,i}), \tag{13}$$

where δ denotes the binary cross-entropy loss function. We optimize the CVR teacher model over the entire space, conditioned on the click propensity $\hat{o}_{u,i}$. the CVR-T loss for EVI, $\mathcal{L}_{\text{CVR-T}}^{\text{EVI}}$, is defined as follows:

$$\mathcal{L}_{\text{CVR-T}}^{\text{EVI}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(r_{u,i}, \hat{r}_{u,i}) | \mathbf{x}_{u,i}, \hat{o}_{u,i}, \tag{14}$$

For the CVR task, we optimize the loss function over the entire exposure space \mathcal{D} and employ inverse click propensity and non-click propensity to ensure the unbiasedness (see Theorem 3):

$$\begin{aligned} \mathcal{L}_{\text{CVR}}^{\text{EVI}} &= \frac{1}{2|\mathcal{D}|} \sum_{(u,i) \in \mathcal{O}} \frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i})}{\hat{o}_{u,i}} \\ &+ \frac{1}{2|\mathcal{D}|} \sum_{(u,i) \in \mathcal{N}} \frac{(1 - o_{u,i}) \delta(r_{u,i}^*, \hat{r}_{u,i})}{1 - \hat{o}_{u,i}} \end{aligned} \tag{15}$$

So the overall loss are defined as:

$$\begin{aligned} \mathcal{L}_{\text{EVI}} &= \lambda_c \mathcal{L}_{\text{CTR}}^{\text{EVI}} + \lambda_t \mathcal{L}_{\text{CVR-T}}^{\text{EVI}} + \lambda_r \mathcal{L}_{\text{CVR}}^{\text{EVI}} \\ &+ \lambda_i \mathcal{L}_{\text{VIE}}^{\text{EVI}} + \lambda_g \mathcal{L}_{\text{CTCVR}}^{\text{EVI}}, \end{aligned} \tag{16}$$

where λ_c , λ_t , λ_r , λ_i and λ_g are trade-offs for each tasks.

5 Experiments

We conduct experiments using five large-scale public datasets and one private industrial dataset to validate the effectiveness of EVI. We compare EVI with 8 CVR estimation baselines. These baselines include both causal and non-causal methods, as well as entire-space and partial-space models. See Appendix A.1 for dataset information, Appendix A.2 for baseline details, and Appendix A.3 for implementation details.

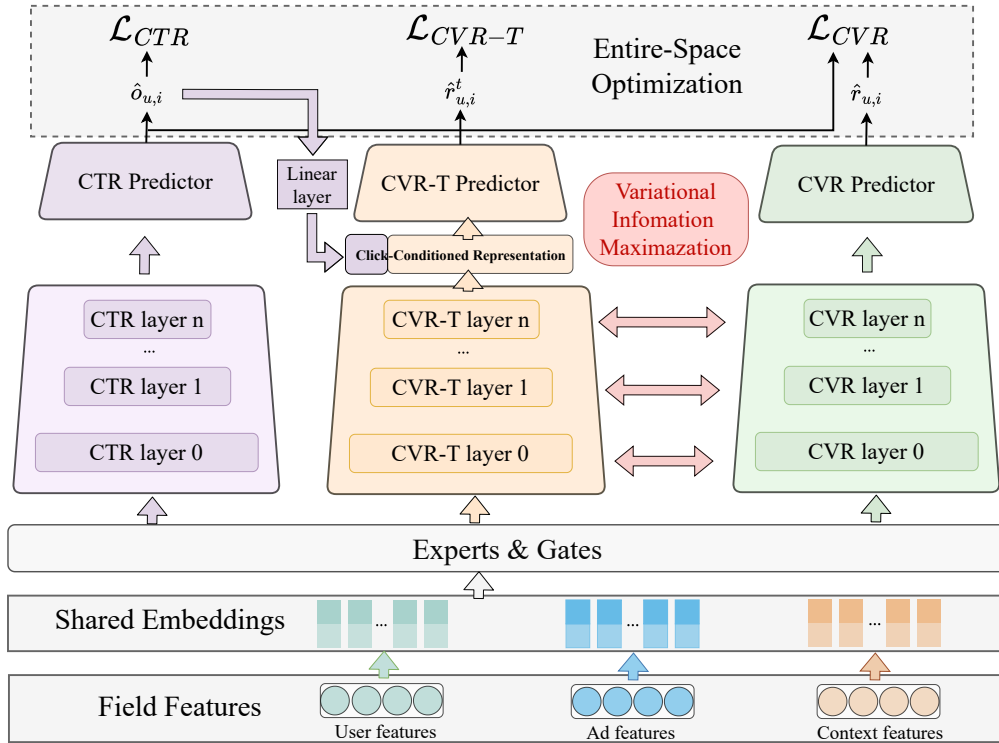


Figure 1: Architecture of EVI. The EVI consists of CTR, CVR-T (teacher) and CVR estimator (student) with shared embedding layer. Following the Experts & Gates, three three multilayer perceptrons serve as representation learners and predictors. CVR-T (teacher) produces unbiased pseudo conversion labels based on click-conditioned representations. We maximize the variational information between CVR-T and CVR representation learner to transfer entire-space conversion knowledge to CVR estimator.

Methods	VIE	CECT	Ali-CCP	AE-FR	AE-ES	AE-US	AE-NL	Industrial
EVI w/o VIE CECT	×	×	0.6702	0.7777	0.7943	0.7888	0.7675	0.8218
EVI w/o VIE	×	✓	0.6776	0.7971	0.8075	0.8015	0.7717	0.8254
EVI	✓	✓	0.6802	0.8008	0.8140	0.8057	0.7741	0.8292

Table 1: Ablation study of EVI. Abbreviations: VIE (variational information exploitation) and CECT (conditional entire-space CVR teacher). Top CVR results are in **bold**.

5.1 Evaluation Metrics

We use the area under the ROC curve (AUC) (Fawcett 2006) and Logloss (NLL) (Xu et al. 2022) as evaluation metrics. Each experiment is conducted five times, and we report the average results. The significance of any improvement or decrease in performance is judged using t-test.

5.2 CVR Estimation Performance Comparison

Table 2 presents a comparison of CVR task results between our EVI model and eight baselines, evaluated across both public and industrial datasets. Key observations can be drawn as:

- The click-space causal methods demonstrate performance comparable to the ESMM model. The ESMM model initially introduced the entire-space CVR estimator. However, it performs poorly on the AE-FR, AE-US,

AE-NL, and industrial dataset. This poor performance may be due to its inherent estimation bias (Wang et al. 2022a). Applying causal inference to correct the inherent bias, click-space causal methods such as MRDR-GPL and DR-V2 outperform ESMM on AE-FR, AE-US, AE-NL, and Industrial datasets. However, they perform worse than ESMM on the Ali-CCP and AE-ES datasets.

- Improved entire-space methods outperform both click-space causal methods and ESMM. Since entire-space exploitation and causal inference are not adversarial, enhanced entire-space methods combine causal inference and entire-space exploitation techniques. The experimental results across all datasets show that improved entire-space methods (EVI, ESCM²-IPW, ESCM²-DR, DCMT, UKD, DDPO, DCMT) outperform both click-space causal methods and ESMM. These findings are consistent with previous studies. Among the enhanced

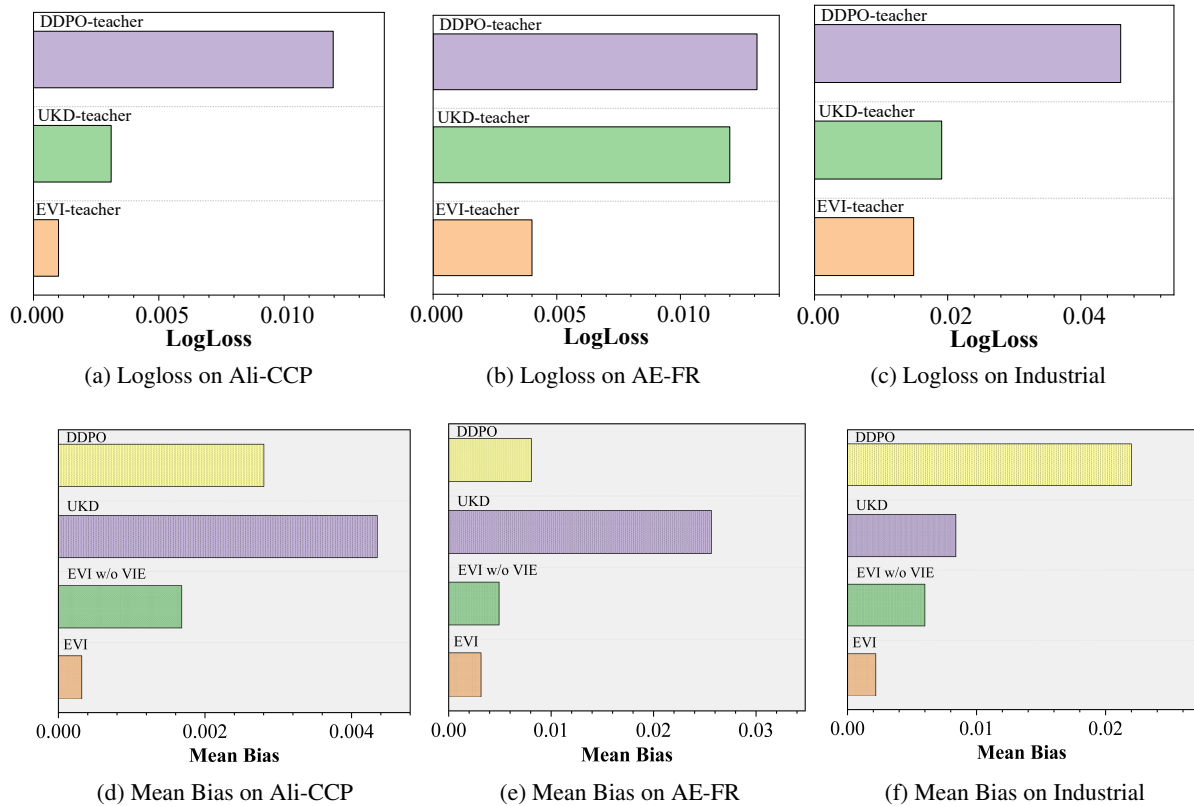


Figure 2: The teachers’ logloss on non-click space and the students’ CVR mean bias. EVI w/o VIE means the EVI excludes variational information exploitation.

entire-space baselines, DDPO performs best in Ali-CCP and AE-NL, while UKD leads in AE-ES and Industrial datasets. ESCM²-IPW and DCMT also excel in AE-FR and AE-US, respectively. Due to varying data characteristics, different methods show varied performance across datasets.

- Our EVI method consistently outperforms all baseline models in CVR estimation AUC across five datasets. It achieves the highest scores in Ali-CCP (0.6802), AE-FR (0.8008), AR-ES (0.8126), AE-US (0.8057), AE-NL (0.7741), and the industrial dataset (0.8292), with improvements of 1.3%, 1.82%, 1.34%, 3.13%, 4.84%, and 0.99% respectively. For NLL, EVI achieves the lowest scores in Ali-CCP (0.0321), AE-FR (0.1037), AR-ES (0.0939), AE-US (0.0976), AE-NL (0.1401), and industrial dataset (0.1621), reducing losses by 1.46%, 1.89%, 1.61%, 2.17%, and 2.77% respectively. In general, EVI shows an average AUC improvement of 2.25% and 2.78% NLL reduction over the optimal baselines.

5.3 Discussion on EVI

In Section 3, through theoretical analysis, we found that biased pseudo labels transfer their bias to the entire-space CVR estimator. Unlike existing methods that use biased CVR teachers, we introduce an unbiased entire-space

teacher trained on a click-conditioned representation. It is crucial to compare the bias of pseudo-labels produced by different CVR teachers using real datasets.

Specifically, we measure the logloss of the pseudo-labels in the non-click space provided by the UKD teacher, DDPO teacher, and EVI teachers. The logloss results conducted on the test sets of Ali-CCP, AE-FR, and the industrial dataset are shown in Figure 2. The logloss for the DDPO-teacher is the highest across the three datasets. This is because the DDPO-teacher is a naïve CVR estimator trained on the click space, making it prone to overestimating the pCVR on the unclick space, consistent with previous studies (Wang et al. 2022a). The logloss of the UKD is lower than that of the DDPO-teacher, possibly due to domain adaptation. Domain adaptation aligns the distributions of the click domain and the non-click domain, thus reducing overestimation. The EVI-teacher attains the lowest logloss, validating the theoretical analysis.

To validate how the pseudo labels and variational information exploitation affect the overall bias of the target CVR estimator, we measure the mean bias of the CVR estimate of the UKD, DDPO, EVI w/o VIE and EVI. The mean bias results, conducted on the test set of Ali-CCP, AE-FR, and the industrial dataset, are shown in Figure 2. Although the pseudo-labels show more bias for DDPO, DDPO shows less

Methods	Ali-CCP		AE-FR		AE-ES		AE-US		AE-NL		Industrial	
	AUC	NLL	AUC	NLL	AUC	NLL	AUC	NLL	AUC	NLL	AUC	NLL
ESMM (Ma et al. 2018b)	0.6703	0.0331	0.7485	0.1148	0.8011	0.0983	0.7241	0.1055	0.6931	0.1463	0.8127	0.1694
ESCM ² -IPW (Wang et al. 2022a)	0.6622	0.0335	<u>0.7865</u>	<u>0.1116</u>	0.7905	0.0962	0.7885	0.1086	0.7554	0.1798	0.8208	0.1644
ESCM ² -DR (Wang et al. 2022a)	0.6494	0.0610	0.7861	0.1269	0.8004	0.1202	0.7887	0.1086	0.7651	0.1755	0.8203	0.1649
MRDR-GPL (Zhou et al. 2023)	0.6195	0.1076	0.7600	0.1972	0.7749	0.1707	0.7624	0.1652	0.7452	0.2312	0.8173	0.2088
DR-V2 (Li et al. 2023b)	0.6093	0.0580	0.7656	0.1192	0.7846	0.1140	0.7591	0.1025	0.7170	0.1592	0.8188	0.2096
UKD (Xu et al. 2022)	0.6594	0.0375	0.7699	0.1156	<u>0.8032</u>	<u>0.0951</u>	0.7790	0.1010	0.7281	0.1448	<u>0.8211</u>	<u>0.1651</u>
DCMT (Zhu et al. 2023)	0.6705	0.0333	0.7718	0.1146	0.7931	0.0952	<u>0.7889</u>	<u>0.0997</u>	0.7339	<u>0.1425</u>	0.8192	0.1684
DDPO (Su et al. 2024)	<u>0.6711</u>	<u>0.0330</u>	0.7751	0.1194	0.7913	0.0976	0.7712	0.1062	<u>0.7685</u>	0.1449	0.8202	0.1687
EVI	0.6802	0.0321	0.8008	0.1037	0.8136	0.0939	0.8057	0.0976	0.7741	0.1401	0.8292	0.1621

Table 2: CVR results for six datasets. Underlined values indicate the best baseline results, and bold values show the top results.

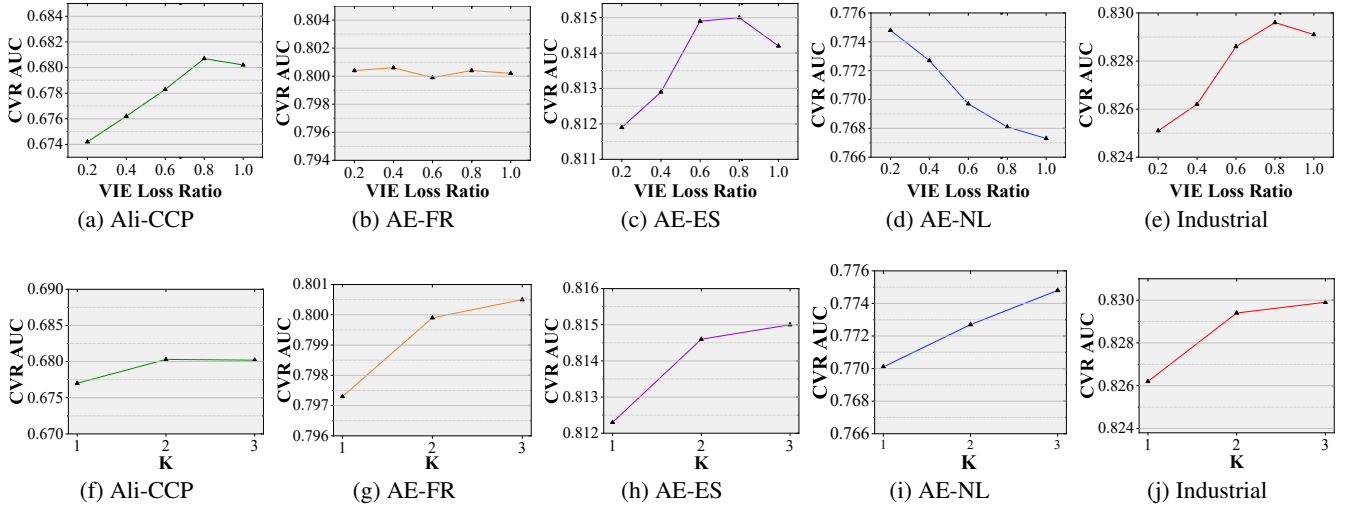


Figure 3: Effects of varying VIE loss ratio and the number of transfer layers K on four public and one industrial datasets.

mean bias on Ali-CCP and AE-FR than UKD. This may be due to the IPS used in DDPO. EVI achieves the lowest mean bias across the three datasets. Specifically, EVI attains a lower mean bias than EVI w/o VIE, indicating that VIE helps EVI to exploit more unclick-space information.

5.4 Ablation Study

The proposed EVI leverages conditional entire-space CVR teacher (CECT) and variational information exploitation (VIE) as its key components. We also perform ablation studies on the CECT and VIE. The ablation study (see Table 1) shows that EVI performs optimally with both CECT and VIE. In six datasets, including only CECT enhances EVI’s performance, but the combination of CECT and VIE achieves the best results. The results highlight the significance of using conditional entire-space CVR teacher and variational information exploitation to improve model performance.

5.5 Parameter Sensitivity Analysis

We conduct a sensitivity analysis on EVI to assess the effects of the variational information exploitation loss ratio and the number of transfer layers, as illustrated in Figure 3. For Ali-CCP, AE-FR, AE-ES, and the industrial dataset,

performance improves as the variational information loss ratio increases up to 1, whereas AE-NL exhibits an opposite trend. Performance improves with more transfer layers as they convey more useful information.

6 Conclusion

This work proposes an entire-space variational information exploitation (EVI) framework to mitigate sample selection bias and data sparsity problems in the CVR estimation task. The EVI ensures unbiased training of the CVR estimator across the entire impression space. To address the bias in pseudo labels of unclicked samples, EVI trains a conditional entire-space CVR teacher model to generate unbiased pseudo labels. Given the limited information carried by pseudo labels, EVI combines logits distillation and variational information lower optimization to enhance the exploitation of non-click space information, thereby improving performance. We conduct extensive experiments on six real-world datasets to validate the effectiveness of our method. In general, EVI shows an average AUC improvement of 2.25% and a 2.78% NLL reduction over the optimal baselines. However, EVI relies on accurate CTR estimation and is sensitive to noisy clicked samples. We’ll explore methods to mitigate this issue.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62176042, in part by TCL Technology Innovation Funding SS2024105, in part by the Fundamental Research Funds for the Central Universities (UESTC) under Grant ZYGX2024Z008, and in part by Tencent Marketing Solution Rhino-Bird Focused Research Program.

References

- Ahn, S.; Hu, S. X.; Damianou, A.; Lawrence, N. D.; and Dai, Z. 2019. Variational information distillation for knowledge transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9163–9171.
- Barber, D.; and Agakov, F. 2004. The im algorithm: a variational approach to information maximization. *Advances in neural information processing systems*, 16(320): 201.
- Dai, Q.; Li, H.; Wu, P.; Dong, Z.; Zhou, X.-H.; Zhang, R.; Zhang, R.; and Sun, J. 2022. A generalized doubly robust learning framework for debiasing post-click conversion rate prediction. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 252–262.
- Fawcett, T. 2006. An introduction to ROC analysis. *Pattern recognition letters*, 27(8): 861–874.
- Feller, W. 1991. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons.
- Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59): 1–35.
- Goldberger; and Greenspan. 2003. An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. In *Proceedings Ninth IEEE International conference on computer vision*, 487–493. IEEE.
- Guo, S.; Zou, L.; Liu, Y.; Ye, W.; Cheng, S.; Wang, S.; Chen, H.; Yin, D.; and Chang, Y. 2021. Enhanced Doubly Robust Learning for Debiasing Post-Click Conversion Rate Estimation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 275–284. Virtual Event Canada: ACM. ISBN 978-1-4503-8037-9.
- Jacobs, R. A.; Jordan, M. I.; Nowlan, S. J.; and Hinton, G. E. 1991. Adaptive mixtures of local experts. *Neural computation*, 3(1): 79–87.
- Li, H.; Xiao, Y.; Zheng, C.; and Wu, P. 2023a. Balancing unobserved confounding with a few unbiased ratings in debiased recommendations. In *Proceedings of the ACM Web Conference 2023*, 1305–1313.
- Li, H.; Xiao, Y.; Zheng, C.; Wu, P.; and Cui, P. 2023b. Propensity matters: Measuring and enhancing balancing for recommendation. In *International Conference on Machine Learning*, 20182–20194. PMLR.
- Luo, H.; Zhuang, F.; Xie, R.; Zhu, H.; Wang, D.; An, Z.; and Xu, Y. 2024. A survey on causal inference for recommendation. *The Innovation*.
- Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; and Chi, E. H. 2018a. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1930–1939. London United Kingdom: ACM. ISBN 978-1-4503-5552-0.
- Ma, X.; Zhao, L.; Huang, G.; Wang, Z.; Hu, Z.; Zhu, X.; and Gai, K. 2018b. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1137–1140.
- Mao, A.; Mohri, M.; and Zhong, Y. 2023. Cross-entropy loss functions: Theoretical analysis and applications. In *International conference on Machine learning*, 23803–23828. PMLR.
- Marlin, B. M.; and Zemel, R. S. 2009. Collaborative prediction and ranking with non-random missing data. In *Proceedings of the third ACM conference on Recommender systems*, 5–12.
- Pearl, J. 2009. Causal inference in statistics: An overview. *Statistics Surveys*.
- Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469): 322–331.
- Saito, Y. 2020. Doubly robust estimator for ranking metrics with post-click conversions. In *Proceedings of the 14th ACM Conference on Recommender Systems*, 92–100.
- Saito, Y.; Yaginuma, S.; Nishino, Y.; Sakata, H.; and Nakata, K. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 501–509.
- Schnabel, T.; Swaminathan, A.; Singh, A.; Chandak, N.; and Joachims, T. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*, 1670–1679. PMLR.
- Su, H.; Meng, L.; Zhu, L.; Lu, K.; and Li, J. 2024. DDPO: Direct Dual Propensity Optimization for Post-Click Conversion Rate Estimation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1179–1188.
- Wang, H.; Chang, T.-W.; Liu, T.; Huang, J.; Chen, Z.; Yu, C.; Li, R.; and Chu, W. 2022a. Escm2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 363–372.
- Wang, Z.; Shen, S.; Wang, Z.; Chen, B.; Chen, X.; and Wen, J.-R. 2022b. Unbiased sequential recommendation with latent confounders. In *Proceedings of the ACM Web Conference 2022*, 2195–2204.
- Wen, H.; Zhang, J.; Lv, F.; Bao, W.; Wang, T.; and Chen, Z. 2021. Hierarchically modeling micro and macro behaviors via multi-task learning for conversion rate prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2187–2191.

Wen, H.; Zhang, J.; Wang, Y.; Lv, F.; Bao, W.; Lin, Q.; and Yang, K. 2020. Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 2377–2386.

Xi, D.; Chen, Z.; Yan, P.; Zhang, Y.; Zhu, Y.; Zhuang, F.; and Chen, Y. 2021. Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising. In *Proceedings of the 27th ACM SIGKDD*, 3745–3755.

Xu, Z.; Wei, P.; Zhang, W.; Liu, S.; Wang, L.; and Zheng, B. 2022. Ukd: Debiasing conversion rate estimation via uncertainty-regularized knowledge distillation. In *Proceedings of the ACM Web Conference 2022*, 2078–2087.

Zhang, D.; Wu, H.; Zeng, G.; Yang, Y.; Qiu, W.; Chen, Y.; and Hu, H. 2022. CTnoCVR: A Novelty Auxiliary Task Making the Lower-CTR-Higher-CVR Upper. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2272–2276.

Zhang, W.; Bao, W.; Liu, X.-Y.; Yang, K.; Lin, Q.; Wen, H.; and Ramezani, R. 2020. Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning. In *Proceedings of The Web Conference 2020*, 2775–2781.

Zheng, H.; Yang, Z.; Liu, W.; Liang, J.; and Li, Y. 2015. Improving deep neural networks using softplus units. In *2015 International joint conference on neural networks (IJCNN)*, 1–4. IEEE.

Zhou, Y.; Feng, T.; Liu, M.; and Zhu, Z. 2023. A Generalized Propensity Learning Framework for Unbiased Post-Click Conversion Rate Estimation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 3554–3563.

Zhu, F.; Zhong, M.; Yang, X.; Li, L.; Yu, L.; Zhang, T.; Zhou, J.; Chen, C.; Wu, F.; Liu, G.; et al. 2023. DCMT: A Direct Entire-Space Causal Multi-Task Framework for Post-Click Conversion Estimation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 3113–3125. IEEE.