# DCMT: A Direct Entire-Space Causal Multi-Task Framework for Post-Click Conversion Estimation

Feng Zhu[1], Mingjie Zhong[1], Xinxing Yang[1], Longfei Li[1], Lu Yu[1], Tiehua Zhang[1], Jun Zhou[1*],
Chaochao Chen[2], Fei Wu[2], Guanfeng Liu[3], and Yan Wang[3]

[1]*Ant Group, Hangzhou, China*
[2]*College of Computer Science and Technology, Zhejiang University, Hangzhou, China*
[3] *School of Computing, Macquarie University, Australia*    * *Corresponding author*

{zhufeng.zhu, mingjie.zmj, xinxing.yangxx, longyao.llf, bruceyu.yl, zhangtiehua.zth, jun.zhoujun}@antgroup.com,
{zjuccc, wufei}@zju.edu.cn, {guanfeng.liu, yan.wang}@mq.edu.au

*Abstract*—In recommendation scenarios, there are two long-standing challenges, i.e., selection bias and data sparsity, which lead to a significant drop in prediction accuracy for both Click-Through Rate (CTR) and post-click Conversion Rate (CVR) tasks. To cope with these issues, existing works emphasize on leveraging Multi-Task Learning (MTL) frameworks (Category 1) or causal debiasing frameworks (Category 2) to incorporate more auxiliary data in the entire exposure/inference space $\mathcal{D}$ or debias the selection bias in the click/training space $\mathcal{O}$. However, these two kinds of solutions cannot effectively address the not-missing-at-random problem and debias the selection bias in $\mathcal{O}$ to fit the inference in $\mathcal{D}$. To fill the research gaps, we propose a Direct entire-space Causal Multi-Task framework, namely DCMT, for post-click conversion prediction in this paper. Specifically, inspired by users' decision process of conversion, we propose a new counterfactual mechanism to debias the selection bias in $\mathcal{D}$, which can predict the factual CVR and the counterfactual CVR under the soft constraint of a counterfactual prior knowledge. Extensive experiments demonstrate that our DCMT can improve the state-of-the-art methods by an average of $1.07\%$ in term of CVR AUC on the offline datasets and $0.75\%$ in term of PV-CVR on the online A/B test (the Alipay Search). Such improvements can increase millions of conversions per week in real industrial applications, e.g., the Alipay Search.

*Index Terms*—casual learning, conversion prediction, and multi-task Learning

## I. INTRODUCTION

Recommender systems (RSs) have been proven to have a powerful filtering capability for users to obtain matched products or services in many industrial applications, e.g., e-commerce platforms [1], [2], search engines [3], payment platforms [4], [5], social networking [6], video-sharing systems [7], and advertising [8], [9]. In these recommender systems, Click-Through Rate (CTR) and post-click Conversion Rate (CVR) predictions are considered as the two most fundamental tasks, which can help online marketplaces understand the underlying logic of their users' click and purchase behaviour. In this paper, we mainly focus on the post-click CVR task.

### A. Challenges

In the above-mentioned recommendation scenarios, users tend to follow a general behaviour path, i.e., "exposure→click
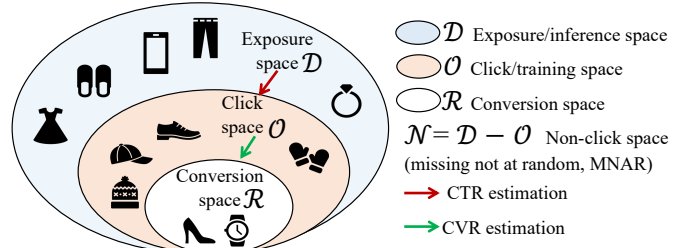


Fig. 1. An example of the data sparsity and selection bias of the CVR estimation task in information-overloaded scenarios, where the training space $\mathcal{O}$ only contains clicked samples, while the inference space is the entire space $\mathcal{D}$ for all exposure samples.

→conversion", meaning both click and conversion labels are recorded in a training sample for RSs [1]. However, for the CVR task, there are two long-standing challenges, i.e., data sparsity and selection bias, which may result in the discrepancy between offline training performance and online testing metrics in real industrial applications.

**Data Sparsity.** Fig. 1 depicts a classical recommendation scenario which contains several exposed samples, a few clicked samples, and very few conversed samples. There are not enough clicked and/or conversed samples for recommender systems to train CTR and/or CVR prediction models. Such a phenomenon is not uncommon in e-commerce recommendation scenarios and referred to as data sparsity. For example, in the benchmark dataset Ali-CCP[1] from Alibaba[2] [1], around $3.75\%$ of exposed items are clicked, and only $0.025\%$ of them are conversed (see the statistic in Table II).

**Selection Bias.** Selection bias is a long-standing issue for conventional post-click CVR models [1], [10], [11]. As shown in Fig. 1, for the conventional post-click CVR models, the data distribution in the training space $\mathcal{O}$ (the click space) is drawn but different from that in target population $\mathcal{D}$ (inference/exposure space). This is because the training data is biased by the user's self-selection. From the perspective of

---

[1]Dataset URL: https://tianchi.aliyun.com/dataset/dataDetail?dataId=408
[2]Alibaba website: https://www.alibaba.com
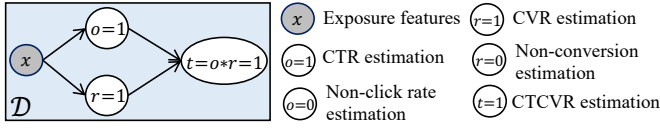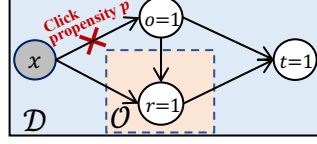
(a) **Parallel MTL approaches** tend to parallelly design CTR estimation task and CVR estimation task and leverage the two auxiliary tasks, i.e., CTR task trained over $\mathcal{D}$ and CTCVR task trained over $\mathcal{D}$, to indirectly predict CVR.
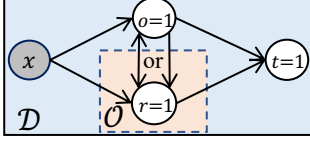


(b) **Multi-gate MTL approaches** tend to transfer the knowledge across the CTR task trained over $\mathcal{D}$ and the CVR task trained over $\mathcal{O}$.

(c) **Propensity-based causal approaches** tend to eliminate the causal effect of input features $x$ (i.e., click propensity $p$) on the CTR task and debias the selection bias in $\mathcal{O}$.
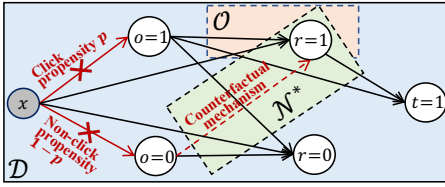


(d) Our **DCMT** aims to eliminate both the causal effect of $x$ (click propensity $p$) on the click event ($o = 1$) and the causal effect of $x$ (non-click propensity $1 - p$) on the non-click event ($o = 0$), which can debias the selection bias in both the factual click space $\mathcal{O}$ and the counterfactual non-click space $\mathcal{N}^*$ (based on a counterfactual mechanism) for the CVR task.

Fig. 2. Existing solutions and our idea.

statistics, the missing data, e.g., exposed&unclicked samples in the non-click space $\mathcal{N}$, is not missing at random (NMAR) [10], [11]. Users tend to click the items they like and thus items with lower CVR are less likely to be observed in $\mathcal{O}$.

### B. Existing Solutions and Their Limitations

**Parallel MTL approaches.** To tackle the above challenges, i.e., data sparsity and selection bias, some of the existing studies (see Fig. 2(a)) tend to treat CTR prediction and CVR prediction as two parallel and related tasks in the same RS and share the input features by using Multi-Task Learning (MTL) frameworks, e.g., ESMM [1] and ESM$^2$ [2]. As shown in Fig. 2(a), for the prediction task of post-click CVR, the MTL approaches [1], [2] tend to share the common embeddings of input features and train CTR and CVR models parallelly. The two auxiliary tasks, i.e., CTR prediction and the click&conversion rate (CTCVR) prediction, are used to indirectly predict CVR via the probabilistic function $p(t = 1|x) = p(o = 1, r = 1|x) = p(o = 1|x) * p(r = 1|x, o = 1)$ [1]. The CTR and CTCVR tasks can be both trained on $\mathcal{D}$, which makes the training space the same as the inference space. This strategy can ease the problems of data sparsity and selection bias to some extent. However, as shown in Fig. 2(a), these existing frameworks actually model another probabilistic function, i.e., $p(t = 1|x) = p(o = 1|x) * p(r = 1|x)$. This means that these approaches ignore that the click event

($o = 1$) has a causal effect on the conversion event ($r = 1$) (**Limitation 1**), and thus fail to take advantage of such underlying correlations. This problem is also called potential independence priority (**PIP**) in the literature [5].

**Multi-gate MTL approaches.** To address **Limitation 1**, the other MTL models, e.g., Cross Stitch [12], MMOE [13], PLE [14], MOSE [15], and AITM [9], are applied to build the relationship between the CTR task trained over $\mathcal{D}$ and the CVR task trained over $\mathcal{O}$. These multi-gate MTL approaches either apply some shared and/or specific expert modules to share the knowledge across the CTR and CVR tasks [12]–[15] or apply a mechanism to transfer the knowledge from the CTR task to the CVR task [9] (see Fig. 2(b)). However, none of these MTL approaches attempt to address the problem of NMAR (**Limitation 2**).

**Causal approaches.** To address **Limitation 2**, some causal approaches with multi-task learning, e.g., Multi-IPW/DR [10] and ESCM$^2$ [5], are proposed to adapt for the data generation process and thus restore the information from NMAR data [10]. As shown in Fig. 2(c), the core idea of these causal approaches is to eliminate the effect of input features $x$ (i.e., click propensity $p$) on the CTR task and thus obtain an conditionally unbiased estimation of CVR $p(r = 1|do(o = 1), x)$. The "do" denotes the do-calculus that are applied to address the confounding bias in causal inference, i.e., the input features $x$ affect both the treatment $o = 1$ and the outcome $r = 1$ [11]. The existing causal approaches mainly focus on debiasing the selection bias in $\mathcal{O}$, e.g., Multi-IPW and ESCM$^2$-IPW, or leveraging a new auxiliary imputation task trained over $\mathcal{D}$ to improve the debiasing performance in $\mathcal{O}$, e.g., Multi-DR and ESCM$^2$-DR. The authors in Multi-DR and ESCM$^2$-DR believe that their strategy can indirectly debias the selection bias in $\mathcal{D}$. However, it is still hard to guarantee that the unbiased CVR estimation trained over $\mathcal{O}$ can infer very well over $\mathcal{D}$ (**Limitation 3**, see the Result 3-2 in Section IV-B2).

### C. Our Approach and Contribution

To further address **Limitation 3**, it is necessary to directly debias the CVR task of the causal models over $\mathcal{D}$ rather than only debias over $\mathcal{O}$, e.g., Multi-IPW and ESCM$^2$-IPW, or indirectly debias over $\mathcal{D}$, e.g., Multi-DR and ESCM$^2$-DR. However, in real application scenarios, there are not positive samples for the CVR task, i.e., the "non-click→conversion" samples, in the non-click space $\mathcal{N}$. These "non-click→conversion" samples are actually hidden in the true negative ("non-click→non-conversion") samples as fake negative samples. $\mathcal{D}$ contains a lot of fake negative samples because users' attitude toward their unclicked items is not entirely negative. It could be that the users have not been aware of these unclicked items because of exposure position, display style, and other factors. Suppose the unclicked items are clicked, it is still possible for the users to purchase the non-click items. This means that blindly utilizing all unclicked samples as negative samples in $\mathcal{N}$ may lead to a discrepancy between users' conversion preference and their

CVR predictions. Instead, in this paper, we propose a **D**irect entire-space **C**ausal **M**ulti-**T**ask framework, namely DCMT, for post-click conversion estimation. Our DCMT designs a new counterfactual mechanism (the underlying rationale will be analysed in Section III-C) to debias the selection bias in the counterfactual non-click space $\mathcal{N}^*$ (a mirror space of $\mathcal{N}$, see Fig. 2(d)). The main contributions of our DCMT framework are summarized as follows.

- This is the first causal work that attempts to estimate the CVR by directly debiasing the selection bias in $\mathcal{D}$ ($\mathcal{O}$ + $\mathcal{N}^*$) rather than only debiasing the selection bias in $\mathcal{O}$. A theoretical proof of the unbiased CVR estimation of our DCMT is provided in this paper.
- Inspired by the decision process of conversion, we propose a new counterfactual mechanism to predict the factual CVR in the factual space and the counterfactual CVR in the counterfactual space under the soft constraint of a counterfactual prior knowledge. With the help of our proposed twin tower, the CVR estimation of our DCMT can effectively simulate users' decision process of conversion.
- We conduct extensive offline experiments on five real-world public datasets and an online A/B test on the Alipay Search system to verify the claims mentioned above. The experiments demonstrate that our DCMT can improve the state-of-the-art methods by an average of $1.07\%$ in term of CVR AUC on the offline datasets and $0.75\%$ in term of PV-CVR on the online A/B test. Such improvements can increase millions of conversions per week in real industrial applications, e.g., the Alipay Search.

## II. PRELIMINARIES

In this section, we first formalize the definitions of CTR, CVR, and CTCVR. Then, to design an MTL framework with a reasonable knowledge transfer potential among these three tasks, we analyse the state-of-the-art MTL frameworks in practice, e.g., ESMM [1]. Finally, to obtain an unbiased CVR estimation over the entire exposure space $\mathcal{D}$, we introduce the core ideas of the state-of-the-art causal MTL approaches, i.e., MTL-IPW estimator and MTL-DR estimator [5], [10]. For the sake of better readability, we list the important notations of this paper in Table I.

### A. Unbiased Estimation

First, if we can obtain the fully observed conversion labels, i.e., the conversion matrix $R$ is fully observed, the ideal loss function of the CVR prediction task (ground truth) can be formulated as:

$$\mathcal{E}^{\text{ground-truth}} = \mathcal{E}(R, \hat{R}) = \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} e(r_{i,j}, \hat{r}_{i,j}), \quad (1)$$

where $e(r_{i,j}, \hat{r}_{i,j})$ is the log loss, i.e., $e(r_{i,j}, \hat{r}_{i,j}) = -r_{i,j}log(\hat{r}_{i,j}) - (1 - r_{i,j})log(1 - \hat{r}_{i,j})$.

However, in real-world applications, we can only observe a part of conversion labels, i.e., $R^{\text{obs}}$, in the click space $\mathcal{O}$, and the others, i.e., $R^{\text{mis}}$, are missing in the non-click space

TABLE I
IMPORTANT NOTATIONS

| Symbol | Definition |
|---|---|
| $\mathcal{U} = \{u_1, ..., u_m\}$ | the set of $m$ users |
| $\mathcal{V} = \{v_1, ..., v_n\}$ | the set of $n$ items |
| $x_{i,j}$ | the input features of a sample (the user-item pair $< u_i, v_j >$), including user features, item features, and other context features |
| $\mathcal{D} = \mathcal{U} \times \mathcal{V}$ | the exposure space, the set of all exposed user-item pairs, $d_{i,j} \in \{0, 1\}$ indicates whether the item $v_j$ is exposed to the user $u_i$ |
| $\mathcal{O} = \{(i,j)\|o_{i,j} = 1, (i,j) \in \mathcal{D}\}; \mathcal{O} \in \mathbb{R}^{m \times n}$ | the click space, $O$ is the click matrix of users on items, $o_{i,j} \in \{0, 1\}$ indicates whether the user $u_i$ click on the item $v_j$ |
| $\mathcal{N} = \{(i,j)\|o_{i,j} = 0, (i,j) \in \mathcal{D}\}; \mathcal{O} \in \mathbb{R}^{m \times n}$ | the non-click space |
| $\mathcal{R} = \{(i,j)\|r_{i,j} = 1, (i,j) \in \mathcal{D}\}; \mathcal{R} \in \mathbb{R}^{m \times n}$ | the conversion space, $R$ is the conversion matrix of users on items, $r_{i,j} \in \{0, 1\}$ indicates whether the user $u_i$ purchases the item $v_j$ |
| $\hat{*}$ | the predicted notations, e.g., $\hat{r}_{i,j}$ represents the predicted CVR of $u_i$ on item $v_j$ |

$\mathcal{N}$. Therefore, naive CVR estimators tend to be trained over $\mathcal{O}$ and their loss function can be formulated as:

$$\begin{aligned} \mathcal{E}^{\text{naive}} &= \mathcal{E}(R^{\text{obs}}, \hat{R}) \\ &= \frac{1}{|\mathcal{O}|} \sum_{(i,j) \in \mathcal{O}} e(r_{i,j}, \hat{r}_{i,j}) \\ &= \frac{1}{|\mathcal{O}|} \sum_{(i,j) \in \mathcal{D}} o_{i,j} e(r_{i,j}, \hat{r}_{i,j}). \end{aligned} \quad (2)$$

The bias of CVR loss between the naive CVR estimators and the ground truth is:

$$\begin{aligned} \text{Bias}^{\text{naive}} &= |\mathcal{E}^{\text{naive}} - \mathcal{E}^{\text{ground-truth}}| \\ &= \left| \frac{1}{|\mathcal{O}|} \sum_{(i,j) \in \mathcal{D}} o_{i,j} e(r_{i,j}, \hat{r}_{i,j}) - \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} e(r_{i,j}, \hat{r}_{i,j}) \right| \end{aligned} \quad (3)$$

Finally, based on the above-mentioned bias of CVR loss, we formalize the definition of unbiased estimation as follows.

**DEFINITION II.1.** *Unbiased Estimation of the CVR Prediction: The CVR estimation of a model $M$ is unbiased when the expectation of the estimated loss in $\mathcal{O}$ equals the loss of the ground truth, i.e., $Bias^M = |E_{\mathcal{O}}(\mathcal{E}^M) - \mathcal{E}^{\text{ground-truth}}| = 0$.*

As introduced in Section I, if data is MNAR, then the data distribution of $\mathcal{O}$ is different from that of $\mathcal{D}$, and thus $\text{Bias}^M \gg 0$ (see Eq.(2)), i.e., the CVR estimation of $M$ is biased.

### B. Multi-Task Learning for CVR Estimation

CTR prediction and CTCVR prediction are auxiliary tasks for the main task, i.e., CVR estimation. The prediction results of CTR and CTCVR can be used to indirectly obtain the prediction result of CVR. This strategy of MTL is widely used in the literature, such as ESMM [1] and ESM$^2$ [2]. The underlying reason behind this strategy is that the traditional CVR prediction task can be only trained on $\mathcal{O}$, i.e., all clicked samples, shown in Fig. 1, and the clicked samples are very sparse. Thus, with the help of the CTR prediction task and CTCVR prediction task (trained on $\mathcal{D}$ with all exposed items shown in Fig. 1), the problems of data sparsity and selection bias are expected to be alleviated to some extent.

However, these parallel MTL approaches cannot obtain an unbiased CVR estimation. We choose the representative method, i.e., ESMM, as an example to demonstrate that the CVR estimation of these parallel MTL approaches is biased, which has been theoretically analysed in the literature [10].

In addition to the parallel MTL methods, e.g., ESMM, many multi-gate MTL methods, e.g., Cross Stitch [12], MMOE [13], PLE [14], MOSE [15], and AITM [9], are proposed to build the relationship between the CTR task and the CVR task. However, none of them attempts to address the problem of NMAR, and thus they cannot guarantee that the CVR estimations derived from their methods are unbiased.

### C. Propensity-based Debiasing for CVR Estimation

As introduced in Section I, the click propensities of users on items, i.e., $P$, leads to the distribution difference between $\mathcal{O}$ and $\mathcal{D}$. Thus, some inverse propensity weighting (IPW)-based approaches, e.g., Multi-IPW [10] and ESCM$^2$-IPW [5], are proposed to eliminate the effect of the input features $x$ on users' click events. The loss function of IPW-based CVR estimators can be formulated as:

$$\mathcal{E}^{\text{IPW}} = \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} \frac{o_{i,j} e(r_{i,j}, \hat{r}_{i,j})}{\hat{p}_{i,j}}, \tag{4}$$

where $\hat{p}_{i,j}$ is the prediction of click propensity of user $u_i$ on item $v_j$. Intuitively, it can eliminate the effect of click propensity on the CVR estimation by giving an inverse propensity weight, i.e., $1/\hat{p}_{i,j}$, to the corresponding CVR loss, i.e., $e(r_{i,j}, \hat{r}_{i,j})$. Note that the click propensity of user $u_i$ on item $v_j$ is the CTR of $u_i$ on item $v_j$, i.e., $\hat{p}_{i,j} = \hat{o}_{i,j}$. Therefore, some existing causal approaches, e.g., Multi-IPW and ESCM$^2$-IPW, leverage multi-task learning to simultaneously learn an auxiliary propensity prediction, i.e., CTR prediction, with CVR prediction. These approaches are categorised into the group of MTL-IPW. The corresponding loss function of MTL-IPW-based CVR estimators can be rewritten as follows.

$$\begin{aligned} \mathcal{E}^{\text{MTL-IPW}} &= \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} \frac{o_{i,j} e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} \\ &= \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}}. \end{aligned} \tag{5}$$

**MTL-IPW Analysis.** The precondition of unbiased estimation of IPW-based approaches is that the prediction of click propensity (CTR prediction) is accurate. However, in practice, it is almost impossible for these IPW-based approaches to guarantee this precondition in the training process. Also, as shown in Eq.(5), the IPW-based CVR approaches only consider the CVR loss in $\mathcal{O}$, i.e., $o_{i,j} = 1$. This means that these IPW-based CVR approaches are actually trained over $\mathcal{O}$ rather than $\mathcal{D}$.

### D. Doubly Robust for CVR Estimation

To address the above-mentioned issues of MTL-IPW approaches, some doubly robust-based approaches, e.g., Multi-DR [10] and ESCM$^2$-DR [5], are proposed in the literature [5], [10]. In addition to CTR task and CVR task, these doubly robust-based approaches also propose an imputation task trained over $\mathcal{D}$ to estimate the CVR error $e_{i,j} = e(r_{i,j}, \hat{r}_{i,j})$ with $\hat{e}_{i,j}$. The performance of this imputation task is assessed by $\delta_{i,j} = e_{i,j} - \hat{e}_{i,j}$. Similarly, these approaches are categorised in the group of MTL-DR. The loss function of MTL-DR-based CVR estimators can be formulated as follows.

$$\mathcal{E}^{\text{MTL-DR}} = \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} \left( \hat{e}_{i,j} + \frac{o_{i,j} \delta_{i,j}}{\hat{o}_{i,j}} \right). \tag{6}$$

**MTL-DR Analysis.** As indicated in Eq. (6), the first term, i.e., $\hat{e}_{i,j}$ (the prediction of CVR error), occurs in $\mathcal{D}$. However, in $\mathcal{N} \in \mathcal{D}$, the conversion label $r_{i,j}$ is not accurate, i.e., there are fake negative samples in this space. The second term, i.e., $\frac{o_{i,j} \delta_{i,j}}{\hat{o}_{i,j}}$ (the weighted loss of the imputation task), only occurs in $\mathcal{O}$. Also, compared with MTL-IPW, the precondition of unbiased estimation of DR-based approaches becomes two 'or' conditions, i.e., the CTR predictions are accurate or the predictions of CVR loss are accurate. However, it is difficult to convince us that the main task, i.e., CVR task, with the two uncertain auxiliary tasks, i.e., CTR task and imputation task, is better than that with the one uncertain auxiliary task, i.e., CTR task.

## III. THE PROPOSED MODEL

All the above-mentioned issues of the existing CVR approaches motivate us to propose our **D**irect entire-space **C**ausal **M**ulti-**T**ask framework, namely DCMT, for post-click conversion estimation scenarios.

### A. Overview of Our Framework

We briefly present each component of DCMT as follows.

**Input (shared features).** First, for the input of our DCMT, we classify the input features into two categories: deep features and wide features. In *CTR Task* and *CVR Task*, we can combine the benefits of generalization (trained on deep features) and memorization (trained on wide features) [16]. In this paper, we collect the features of user profiles and item details, e.g., gender (user), age (user), and shop id (item), as the deep features, which can be trained to learn the deep and complex relationship between users and items. We collect the interaction features, e.g., favourite shop id (from user to item), which can be trained to understand users' historical preferences. Note that if a training dataset does not contain any wide features, our DCMT framework will degenerate from a wide&deep structure to a pure deep structure.

**Embedding Layer.** In this layer, the input features are embedded into low-dimension spaces and we concatenate the embeddings of deep features, and wide features, respectively. The concatenated deep embeddings and wide embeddings are shared by *CTR Task* and *CVR Task*.

**CTR Task.** In this component, we adopt a Multi-Layer Perceptron (MLP), i.e., CTR tower (deep part) in Fig. 3, to represent the deep relation between users and items. Also, we adopt a generalized linear structure, i.e., CTR tower (wide part),
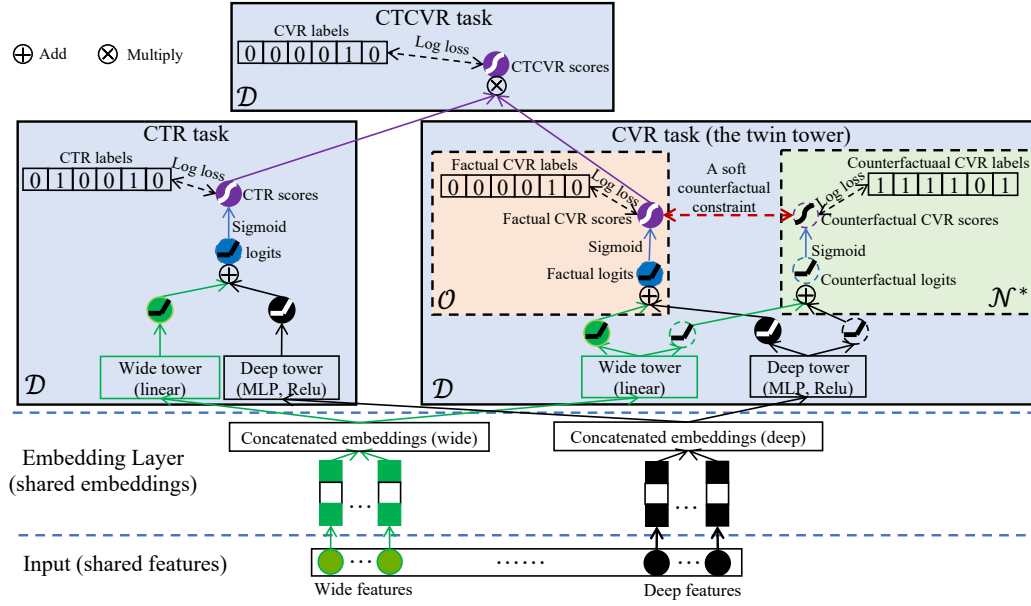
Fig. 3. The structure of our DCMT model.

which can be trained to capture users' historical preferences over $\mathcal{D}$. Based on these two CTR towers, we can obtain their corresponding outputs, i.e., CTR predictions.

**CVR Task.** Like *CTR Task*, the wide and deep concatenated embeddings are fed into the twin tower of *CTR Task*. In this task, we will propose a new counterfactual mechanism to predict the factual CVR and the counterfactual CVR simultaneously. These two CVR predictions are generated by the twin tower (see its structure in Fig. 3) that will be introduced in Section III-E. The two predictions follow a counterfactual prior knowledge, i,e, the sum of them should be 1. With the help of the counterfactual mechanism, our CVR estimator can leverage all the samples in $\mathcal{D}$, which can address the problem of data sparsity. Compared with the traditional propensity-based debiasing approaches, our counterfactual mechanism can also help our DCMT debias the selection bias in the counterfactual non-click space $\mathcal{N}^*$. Note that the counterfactual CVR task is an auxiliary task for the factual CVR task, and we only retain the predictions of the factual CVR task as the final results of the CVR task.

**CTCVR Task.** Finally, we can obtain the CTCVR predictions over $\mathcal{D}$ by multiplying the CTR predictions and the CVR predictions, i.e., $\hat{t}_{i,j} = \hat{o}_{i,j} * \hat{r}_{i,j}$.

Since the CVR task is our main task, and thus, we will introduce the details of *CVR Task* in the following sections.

### B. CVR Estimation over the Entire Exposure Space $\mathcal{D}$

As introduced in Section II-C, the IPW-based approaches focus on debiasing the selection bias caused by click propensity in $\mathcal{O}$. However, these approaches seem to ignore the samples in $\mathcal{N}$ in the training of CVR models. To leverage all the samples in $\mathcal{D}$ and alleviate the problems of selection bias and data

sparsity in $\mathcal{D}$, the CVR loss function of naive solutions is formulated as follows.

$$
\begin{aligned}
\mathcal{E}^{\text{DCMT\_naive}} &= \frac{1}{|\mathcal{D}|} \left( \sum_{(i,j)\in\mathcal{O}} \frac{e_{i,j}}{\hat{o}_{i,j}} + \sum_{(i,j)\in\mathcal{N}} \frac{e_{i,j}}{1-\hat{o}_{i,j}} \right) \\
&= \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} \left( \frac{o_{i,j}e_{i,j}}{\hat{o}_{i,j}} + \frac{(1-o_{i,j})e_{i,j}}{1-\hat{o}_{i,j}} \right),
\end{aligned}
\tag{7}
$$

where $e_{i,j} = e(r_{i,j}, \hat{r}_{i,j})$. In contrast to the click propensity, i.e., $\hat{p}_{i,j}$, in $\mathcal{O}$, the non-click propensity, i.e., $1 - \hat{p}_{i,j}$, in $\mathcal{N}$ also leads to a selection bias. However, in $\mathcal{N}$, the conversion labels are always 0, i.e., $r_{i,j} = 0, \forall(i,j) \in \mathcal{N}$. There are fake negative samples in $\mathcal{N}$. This is the main reason that the conventional CVR estimators did not leverage the samples in $\mathcal{N}$ to reduce the selection bias. Therefore, we cannot directly apply Eq. (7) to reduce the selection bias caused by click propensity in $\mathcal{D}$.

### C. Counterfactual Mechanism

As introduced in Section II-D, the DR-based approaches apply an auxiliary imputation task to predict the CVR error, i.e., $\hat{e}_{i,j}$. This imputation tower is trained over $\mathcal{D}$, and thus it is expected to reduce the selection bias in $\mathcal{D}$. However, we have analysed its disadvantages in Section II-D. To address these disadvantages, in this section, we propose another promising solution, i.e., a new counterfactual mechanism, to achieve the same goal of training over $\mathcal{D}$.

**Underlying Rationale.** Inspired by the decision process of conversion, we propose a concept of counterfactual sample space for CVR tasks (see Fig. 4). From the perspective of probability, the user $u_i$ purchases the item $v_j$ with a probability of $p_{i,j}$ (conversion rate), and thus $u_i$ does not purchase $v_j$ with a probability of $1 - p_{i,j}$. Suppose the decision of conversion is true, i.e., $r_{i,j} = 1$, then there is a positive sample $< r_{i,j} = 1, o_{i,j} = 1, x_{i,j} >$ in the factual exposure $\mathcal{D}$.
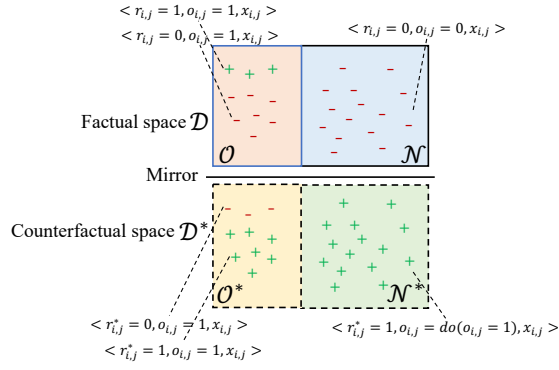
Fig. 4. The counterfactual space is the mirror space of the corresponding factual space. A counterfactual sample is its corresponding factual sample with an opposite status (label).
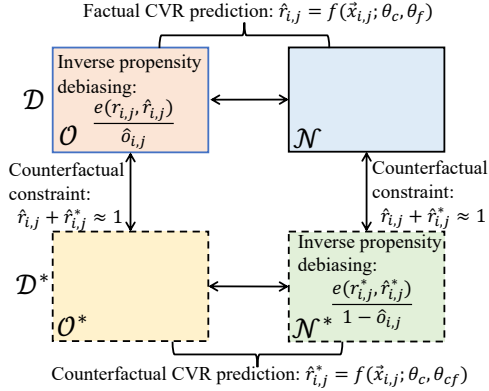


Fig. 5. Debiasing circle. The debiasing strategies, i.e., $\frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}}$ in $\mathcal{O}$ and $\frac{e(r^*_{i,j}, \hat{r}^*_{i,j})}{1-\hat{o}_{i,j}}$ in $\mathcal{N}^*$, can indirectly affect the CVR predictions in $\mathcal{O}^*$ and $\mathcal{N}$ respectively, via the soft counterfactual constraint, i.e., $\hat{r}_{i,j} + \hat{r}^*_{i,j} \approx 1$.

However, in this case, we cannot observe a negative sample $< r_{i,j} = 0, o_{i,j} = 1, x_{i,j} >$ though this counterfactual negative sample could have been observed with the probability of $1 - p_{i,j}$. In our setting of counterfactual mechanism, this counterfactual negative sample will be included in the counterfactual exposure space $\mathcal{D}^*$. From the perspective of physics, in this paper, the counterfactual space $\mathcal{D}^*$ is the space mirror of the factual space $\mathcal{D}$. A counterfactual sample is its corresponding factual sample with an opposite status, which is similar to the relation between matter and antimatter [17].

**A Different Counterfactual Mechanism.** The traditional counterfactual mechanism in causal inference tends to extend the factual/observed sample space. These counterfactual samples may have the same or similar features as the observed samples and their labels are predicted or generated by reasonable prior knowledge. However, this mechanism still cannot guarantee that the predicted/generated labels are accurate or the distribution of counterfactual samples is unbiased.

Since the definition of counterfactual samples in this paper is different from that in traditional causal inference [18], for readability, we formalize the definition of counterfactual samples as follows.

**DEFINITION III.1.** *Counterfactual Samples*: A counter-

factual sample is the mirror image of the corresponding factual sample, e.g., $< r^*_{i,j} = 1 - r_{i,j}, o_{i,j}, x_{i,j} >$ vs. $< r_{i,j}, o_{i,j}, x_{i,j} >$.

As shown in Fig. 4, for a unclicked&unconversed sample $< r_{i,j} = 0, o_{i,j} = 0, x_{i,j} >$ in the non-click space $\mathcal{N}$, we can generate its corresponding counterfactual sample $< r^*_{i,j} = 1, do(o_{i,j}) = 1, x_{i,j} >$ in $\mathcal{N}^*$. Here, the "do" denotes the do-calculus that are applied to address the confounding bias in causal inference. In this case, "do" means that we suppose the unclicked&unconversed sample is clicked, then we can predict its counterfactual CVR based on its corresponding counterfactual sample.

Therefore, under the setting of our counterfactual mechanism, the CVR loss function of our DCMT can be rewritten as follows.

$$
\begin{aligned}
\mathcal{E}^{\text{DCMT\_main}} \\
= \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} \left( \frac{o_{i,j} e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + \frac{(1-o_{i,j}) e(1-r_{i,j}, \hat{r}^*_{i,j})}{1-\hat{o}_{i,j}} \right) \\
= \frac{1}{|\mathcal{D}|} \left( \sum_{(i,j)\in\mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + \sum_{(i,j)\in\mathcal{N}} \frac{e(1-r_{i,j}, \hat{r}^*_{i,j})}{1-\hat{o}_{i,j}} \right) \\
= \frac{1}{|\mathcal{D}|} \left( \sum_{(i,j)\in\mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + \sum_{(i,j)\in\mathcal{N}^*} \frac{e(r^*_{i,j}, \hat{r}^*_{i,j})}{1-\hat{o}_{i,j}} \right),
\end{aligned}
\tag{8}
$$

where $\hat{r}^*_{i,j}$ is the corresponding prediction of counterfactual CVR (see Eq. (12)). Actually, in Eq. (8), the second term suffers from the problem of fake negative samples as well because we only reverse the conversion status of samples from $\mathcal{N}$ to $\mathcal{N}^*$. To address this issue, we employ a counterfactual regularizer in the following section.

**Counterfactual Prior Knowledge and Regularizer.** As mentioned in **Underlying Rationale**, the user $u_i$ purchases the item $v_j$ with a probability of $p_{i,j}$ (conversion rate), and thus $u_i$ does not purchase $v_j$ with a probability of $1 - p_{i,j}$. The prediction of factual CVR $\hat{r}_{i,j}$ and the prediction of counterfactual CVR $\hat{r}^*_{i,j}$ should also follow this prior knowledge, i.e., $\hat{r}_{i,j} + \hat{r}^*_{i,j} = 1$. However, if we force $\hat{r}_{i,j} + \hat{r}^*_{i,j}$ to be 1, then $\hat{r}_{i,j}$ and $\hat{r}_{i,j}$ will be restricted in respective small value ranges, e.g.,$[0.265, 0.305]$ and $[0.695, 0.735]$ (see Figure 8(d) in Section IV-C). This kind of hard constraint makes the training process of our CVR estimation unable to minimize the factual loss and the counterfactual loss in Eq. (8). Therefore, the hard constraint is harmful to our CVR estimation. Instead, we add a counterfactual regularizer $L$ (a soft constraint) in Eq. (8) to minimize the error between 1 and $\hat{r}_{i,j} + \hat{r}^*_{i,j}$ as follows.

$$
\begin{aligned}
\mathcal{E}^{\text{DCMT}} = \mathcal{E}^{\text{DCMT\_main}} + L \\
= \frac{1}{|\mathcal{D}|} \left( \underbrace{\sum_{(i,j)\in\mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}}}_{\text{factual loss in } \mathcal{O}} + \underbrace{\sum_{(i,j)\in\mathcal{N}^*} \frac{e(r^*_{i,j}, \hat{r}^*_{i,j})}{1-\hat{o}_{i,j}}}_{\text{counterfactual loss in } \mathcal{N}^*} \right) \\
+ \underbrace{\frac{\lambda_1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} \left| 1 - (\hat{r}_{i,j} + \hat{r}^*_{i,j}) \right|}_{\text{counterfactual regularizer } L},
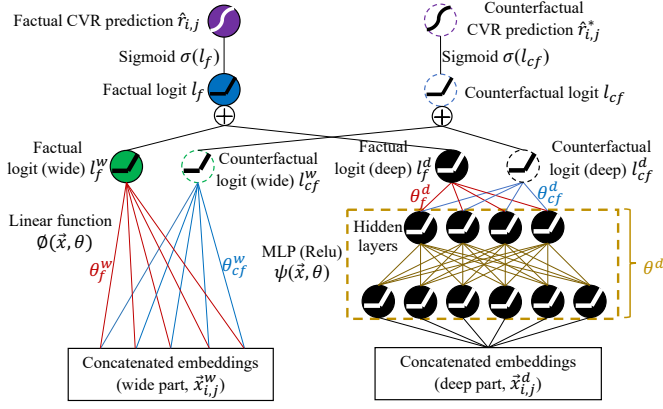\end{aligned}
\tag{9}
$$

Fig. 6. The detailed wide&deep structure of our twin tower. In contrast to the general prediction function of our twin tower, i.e., Eq. (11), $\theta_c = \theta^d$ (there are only the common parameters $\theta^d$ in the deep part), $\theta_f = \theta_f^w + \theta_f^d$, and $\theta_{cf} = \theta_{cf}^w + \theta_{cf}^d$. The detailed function is formulated in Eq. (12).

where $\lambda_1$ is the hyper-parameter to control the importance of the counterfactual regularizer. The counterfactual regularizer can help to eliminate the negative effect of the fake negative samples in $\mathcal{N}$, i.e., the fake positive samples in $\mathcal{N}^*$. As indicated in Eq. (9), the first term, i.e., factual loss in $\mathcal{O}$, is used to debias the selection bias in $\mathcal{O}$, while the second term, i.e., counterfactual loss in $\mathcal{N}^*$, is used to debias the selection bias in $\mathcal{N}^*$.

With the help of the counterfactual regularizer, i.e., the third term in Eq. (9), the CVR task of our DCMT actually forms an interesting debiasing circle as shown in Fig. 5. The debiasing strategy, i.e., $\frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}}$, in $\mathcal{O}$ directly affects the predictions of factual CVR in $\mathcal{N}$ via the shared prediction function, i.e., $\hat{r}_{i,j} = f(\vec{x}_{i,j}; \theta_c, \theta_f)$, and indirectly affects the predictions of counterfactual CVR in $\mathcal{O}^*$ via the soft counterfactual constraint, i.e., $\hat{r}_{i,j} + \hat{r}_{i,j}^* \approx 1$. Similarly, the debiasing strategy, i.e., $\frac{e(r_{i,j}^*, \hat{r}_{i,j}^*)}{1-\hat{o}_{i,j}}$, in $\mathcal{N}^*$ directly affects the predictions of counterfactual CVR in $\mathcal{O}^*$ via the shared prediction function, i.e., $\hat{r}_{i,j}^* = f(\vec{x}_{i,j}; \theta_c, \theta_{cf})$, and indirectly affects the predictions of factual CVR in $\mathcal{N}$. This debiasing circle can achieve the goal of debiasing the selection bias in both $\mathcal{D}$ and $\mathcal{D}^*$, which can address the disadvantages of MTL-IPW and MTL-DR. The prediction functions of the factual CVR and the counterfactual CVR will be carefully introduced in the next section.

### D. Unbiasd CVR Estimation of Our DCMT

**THEOREM III.1.** *The CVR estimation of our DCMT is unbiased in the exposure space $\mathcal{D}$, i.e., $Bias^{DCMT} = |E_{\mathcal{O}}(\mathcal{E}^{DCMT}) - \mathcal{E}^{ground\text{-}truth}| = 0$, when $o_{i,j} = \hat{o}_{i,j}$ and $\hat{r}_{i,j} + \hat{r}_{i,j}^* = 1$, i.e., the prediction of click propensity is accurate and the predictions of factual CVR and counterfactual CVR follow the counterfactual prior knowledge.*

Note that $o_{i,j} = \hat{o}_{i,j}$ means that $o_{i,j} = \hat{o}_{i,j} = 1$ in the click space $\mathcal{O}$ and $o_{i,j} = \hat{o}_{i,j} = 0$ in the non-click space $\mathcal{N}$. Meanwhile, $\hat{r}_{i,j} + \hat{r}_{i,j}^* = 1$ means that the counterfactual regularizer $L = 0$ and $e(1 - r_{i,j}, \hat{r}_{i,j}^*) = e(1 - r_{i,j}, 1 - \hat{r}_{i,j}) = e(r_{i,j}, \hat{r}_{i,j})$ because $e(r_{i,j}, \hat{r}_{i,j})$ is the log loss.

*Proof.*

$$
Bias^{DCMT} = \left| \mathcal{E}^{DCMT} - \mathcal{E}^{ground\text{-}truth} \right|
$$

$$
= \left| \frac{1}{|\mathcal{D}|} \left( \sum_{(i,j)\in\mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + \sum_{(i,j)\in\mathcal{N}^*} \frac{e(r_{i,j}^*, \hat{r}_{i,j}^*)}{1-\hat{o}_{i,j}} \right) \right.
$$
$$
\left. + L - \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} e(r_{i,j}, \hat{r}_{i,j}) \right|
$$

$$
= \left| \frac{1}{|\mathcal{D}|} \left( \sum_{(i,j)\in\mathcal{O}} \frac{e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + \sum_{(i,j)\in\mathcal{N}} \frac{e(1 - r_{i,j}, \hat{r}_{i,j}^*)}{1-\hat{o}_{i,j}} \right) + L \right.
$$
$$
\left. - \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} e(r_{i,j}, \hat{r}_{i,j}) \right|
$$

$$
= \left| \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{O}} \frac{(1 - \hat{o}_{i,j})e(r_{i,j}, \hat{r}_{i,j})}{\hat{o}_{i,j}} + L \right.
$$
$$
\left. + \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{N}} \frac{e(1 - r_{i,j}, \hat{r}_{i,j}^*) - (1 - \hat{o}_{i,j})e(r_{i,j}, \hat{r}_{i,j})}{1-\hat{o}_{i,j}} \right| = 0.
$$

$$(10)$$

$\square$

### E. Twin Tower

As introduced in **Underlying Rationale** of Section III-C, our counterfactual mechanism is inspired by the decision process of conversion. To simulate users' decision process of conversion, we design a twin tower to predict the factual CVR and the counterfactual CVR simultaneously with the same input features. The general prediction function of the twin tower is formulated as follows.

$$
< \hat{r}_{i,j}, \hat{r}_{i,j}^* >= f(\vec{x}_{i,j}; \theta_c, \theta_f, \theta_{cf}), \tag{11}
$$

where $\vec{x}_{i,j}$ is the feature embedding vector of the input features $x_{i,j}$, $\theta_c$ are the common parameters for both $\hat{r}_{i,j}$ and $\hat{r}_{i,j}^*$, $\theta_f$ are the specific parameters for $\hat{r}_{i,j}$, and $\theta_{cf}$ are the specific parameters for $\hat{r}_{i,j}^*$. The common parameters represent the same thoughts for decision making when receiving the same input, while the specific parameters represent the divergent thoughts to make the final decisions, i.e., conversion or non-conversion. In the training process, the common parameters $\theta_c$ are updated by minimizing both the losses of the factual CVR task and the counterfactual CVR task. In contrast to existing CVR estimators, e.g., MTL-IPW and MTL-DR, this strategy may avoid that the parameters of CVR estimators are biased to the samples in $\mathcal{D}$, especially for $\mathcal{O}$.

Since we adopt a wide&deep structure for our CVR estimator, as shown in Fig. 6, the prediction function of our twin tower can be specifically formulated as follows.

$$
\begin{aligned}
< \hat{r}_{i,j}, \hat{r}_{i,j}^* > &= < \sigma(l_f), \sigma(l_{cf}) > \\
&= < \sigma(l_f^w + l_f^d), \sigma(l_{cf}^w + l_{cf}^d) > \\
&= < \sigma(\phi(\vec{x}_{i,j}^w; \theta_f^w) + \psi(\vec{x}_{i,j}^d; \theta^d, \theta_f^d)), \\
&\quad \sigma(\phi(\vec{x}_{i,j}^w; \theta_{cf}^w) + \psi(\vec{x}_{i,j}^d; \theta^d, \theta_{cf}^d)) >,
\end{aligned} \tag{12}
$$

where $\sigma(*)$ is the *Sigmoid* function, $\phi(\vec{x}, \theta)$ is the *Linear Regression* function with the input embedding $\vec{x}$ and the parameters $\theta$, and $\psi(\vec{x}, \theta)$ is the *Multi-Layer Perceptron (MLP)* function with the input embedding $\vec{x}$ and the parameters $\theta$. In

addition, $l_f$ is the factual CVR logit, $l_{cf}$ is the counterfactual CVR logit, $l^w$ (e.g., $l_f^w$ and $l_{cf}^w$) is the CVR logit in the wide part, $l^d$ (e.g., $l_f^d$ and $l_{cf}^d$) is the CVR logit in the deep part, $\vec{x}_{i,j}^w$ is the wide feature embedding vector of the user-item pair $<u_i, v_j>$, $\vec{x}_{i,j}^d$ is the deep feature embedding vector of the user-item pair $<u_i, v_j>$, $\theta_f^w$ is the parameters of $\phi$ for $l_f^w$ in the wide part, $\theta_{cf}^w$ are the parameters of $\phi$ for $l_f^w$ in the wide part, $\theta^d$ are the common parameters of $\psi$ in the deep part, $\theta_f^d$ are the specific parameters of $\psi$ for $l_f^d$ in the deep part, and $\theta_{cf}^d$ are the specific parameters of $\psi$ for $l_{cf}^d$ in the deep part. For readability, we carefully mark these notations in the structure of the twin tower (see Fig. 6).

### F. Self-Normalization

To reduce the variance of IPW-based approaches, the Self-Normalized Inverse Propensity Scoring (SNIPS) estimator is widely used in the literature [19], [20]. We also adopt this self-normalization strategy to reduce the variance of our DCMT estimator. Finally, the inverse propensity weights $\frac{1}{\hat{o}_{i,j}}$ and $\frac{1}{(1-\hat{o}_{i,j})}$ in Eq. (9) will be replaced with the following two self-normalized weights.

$$\frac{\frac{1}{\hat{o}_{i,j}}}{\sum_{(i,j)\in\mathcal{O}}\frac{1}{\hat{o}_{i,j}}} \quad \text{and} \quad \frac{\frac{1}{1-\hat{o}_{i,j}}}{\sum_{(i,j)\in\mathcal{N}^*}\frac{1}{1-\hat{o}_{i,j}}}. \quad (13)$$

Also, to avoid the Nan loss in the training process, we need to clip the value range of $\hat{o}_{i,j}$ from the default range, i.e., $[0,1]$, to $(0,1)$.

### G. Training Loss

As mentioned in Section I, we use the two auxiliary tasks, i.e., *CTR Task* and *CTCVR Task*, to help to obtain the predictions of the target *CVR Task*. Therefore, the training loss of our DCMT framework contains three parts, i.e., the loss of *CTR Task*, the loss of *CVR Task*, and the loss of *CTCVR Task*, as follows:

$$L(\theta) = \mathcal{E}^{\text{CTR}} + w^{cvr}\mathcal{E}^{\text{DCMT}} + w^{ctcvr}\mathcal{E}^{\text{CTCVR}} + \lambda_2\|\theta\|_F^2, \quad (14)$$

where $\mathcal{E}^{\text{DCMT}}$ is the CVR loss of our DCMT framework (see Eq. (9)), $\theta$ denotes all network weights in our DCMT framework, $\|\theta\|_F^2$ is the regularizer, $\lambda_2$ is a hyper-parameter that controls the importance of the regularizer. In addition, $w^{cvr}$, $w^{ctcvr}$ are loss weights of $\mathcal{E}^{\text{DCMT}}$, $\mathcal{E}^{\text{CTCVR}}$, which are set to 1 in this paper, respectively. The losses of the CTR task and the CTCVR task can be represented as follows.

$$\mathcal{E}^{\text{CTR}} = \frac{1}{|\mathcal{D}|}\sum_{(i,j)\in\mathcal{D}} e(o_{i,j}, \hat{o}_{i,j}),$$
$$\mathcal{E}^{\text{CTCVR}} = \frac{1}{|\mathcal{D}|}\sum_{(i,j)\in\mathcal{D}} e(r_{i,j}, \hat{t}_{i,j}). \quad (15)$$

TABLE II
EXPERIMENTAL DATASETS

| Dataset | | #User | #Item | Split | #Exposure | #Click | #Conversion |
|---|---|---|---|---|---|---|---|
| **Public (offline test)** | **Ali-CCP** | 0.4M | 4.3M | Train | 42.3M | 1.6M | 9,K |
| | | | | Test | 43M | 1.7M | 9.4K |
| | **AE-ES** | 0.6M | 1.4M | Train | 22.3M | 0.57M | 12.9K |
| | | | | Test | 9.3M | 0.27M | 6.1K |
| | **AE-FR** | 0.57M | 1.2M | Train | 18.2M | 0.34M | 9K |
| | | | | Test | 8.8M | 0.2M | 5.3K |
| | **AE-NL** | 0.37M | 0.81M | Train | 12.2M | 0.25M | 8.9K |
| | | | | Test | 5.6M | 0.14M | 4.9K |
| | **AE-US** | 0.5M | 1.3M | Train | 20M | 0.29M | 7K |
| | | | | Test | 7.5M | 0.16M | 3.9K |
| **Industrial (online test)** | **Alipay Search** | 73M | 531K | Train | 665M | 118M | 88M |
| | | | | Test | 162M | 29M | 22M |

## IV. EXPERIMENTS AND ANALYSIS

We conduct extensive experiments on real-world benchmark datasets and online systems to answer the following key questions:

- **Q1** How does our DCMT model perform when compared with the state-of-the-art (SOTA) models on offline datasets (see Result 1)?
- **Q2** How much does the counterfactual mechanism contribute to performance improvement on offline datasets (see Result 2)?
- **Q3** How does our DCMT model perform when compared with the base model in online environments (see Result 3)?
- **Q4** How do the hyper-parameters affect the performance of our DCMT model (see Result 4)?

### A. Experimental Settings

We conduct extensive experiments on both the offline datasets collected from real-world e-commerce & express systems, and online searching environments.

*1) Experimental Datasets:* To validate the prediction performance of our DCMT model and baseline models, we choose two real-world benchmark datasets, i.e., Ali-CCP[3] (Alibaba Click and Conversion Prediction) [1] and Ali-Express (AE)[4] [21]. Since Ali-Express includes the real-world traffic logs of the Ali-Express search system from different countries, we choose four suitable sub-datasets from four countries, i.e., Spain (AE-ES), French (AE-FR), Netherlands (AE-NL), and America (AE-US). Both datasets encompass user features, item features, combination features (only in Ali-CCP), context features (only in Ali-CCP), and labels of click and conversion. Also, we validate the performance of our DCMT on our Alipay Search dataset (industrial dataset). The Alipay Search dataset comes from a 10-day offline log of the service searching of our Alipay platform, divided into training (7 days), validation (1 day), and testing (2 days) in chronological order. The Alipay Search dataset includes a random sample of the users' searching historical data on the services provided by our Alipay platform. We treat the click on the detailed pages of services, i.e., the second click, as conversion. To protect

---

[3]Dataset URL: https://tianchi.aliyun.com/dataset/dataDetail?dataId=408
[4]Dataset URL: https://tianchi.aliyun.com/dataset/dataDetail?dataId=74690

TABLE III

THE COMPARISON OF THE BASELINES AND OUR METHODS

| Model | | Structure | Main Ideas |
|---|---|---|---|
| Baselines | Parallel MTL Baselines — **ESMM** [1] | Shared bottom | Feature representation transfer learning |
| | Multi-gate MTL baselines — **Cross Stitch** [12] | Cross-stitch unit | Activation combination |
| | **MMOE** [13] | Gated mixture-of-experts | Trade-offs between task-specific objectives and inter-task relationships |
| | **PLE** [14] | Customized gates & local experts & shared experts | Customized sharing (avoiding negative transfer) |
| | **AITM** [9] | Shared bottom & inter-task transfer | Adaptive information transfer |
| | Causal baselines — **ESCM$^2$-IPW** [5] | Two towers (CTR+CVR) | Propensity-based debiasing |
| | **ESCM$^2$-DR** [5] | Three towers (CTR+CVR+Imputation) | Propensity-based debiasing & doubly robust estimation |
| Our methods | Simplified versions — **DCMT_PD** | CTR tower + the twin CVR tower | Propensity-based debiasing over $\mathcal{D}$ |
| | **DCMT_CF** | CTR tower + the twin CVR tower | Counterfactual mechanism |
| | Completed version — **DCMT** | CTR tower + the twin CVR tower | Propensity-based debiasing & counterfactual mechanism |

TABLE IV

THE OFFLINE EXPERIMENTAL RESULTS (AUC) IN DIFFERENT PUBLIC DATASETS FOR *CVR Task* AND *CTCVR Task* (THE BEST-PERFORMING BASELINES WITH RESULTS MARKED WITH * WHILE OUR BEST-PERFORMING MODELS WITH RESULTS MARKED WITH BLACK BODY)

| Dataset | Parallel MTL Baselines ESMM | | Multi-gate MTL Baselines | | | | | | | | Causal Baselines | | | | Our Models (for ablation study) | | | | | | Improvement (DCMT vs. best-performing baselines) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Cross Stitch | | MMOE | | PLE | | AITM | | ESCM$^2$-IPW | | ESCM$^2$-DR | | DCMT_PD | | DCMT_CF | | DCMT | | | |
| | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR | CVR | CTCVR |
| Ali-CCP | .6291 | .6243* | .5637 | .5584 | .6041 | .5947 | .5871 | .5599 | .6324* | .6121 | .6156 | .5932 | .5914 | .6195 | .6352 | .6156 | .6366 | .6141 | **.6486** | **.6341** | 2.56%↑ | 1.57%↑ |
| AE-ES | .6293 | .6027 | .6855 | .6453 | .6082 | .6211 | .5262 | .6146 | .7109 | .7044 | .8561 | .8688 | .8601* | .8681* | .8422 | .8329 | .8612 | .8557 | **.8677** | **.8817** | 0.88%↑ | 1.57%↑ |
| AE-FR | .6175 | .6039 | .7186 | .6443 | .6290 | .0.6451 | .5827 | .5351 | .7316 | .6923 | .8524* | .8647 | .8513 | .8668* | .8087 | .8512 | .8271 | .8470 | **.8576** | **.8756** | 0.61%↑ | 1.02%↑ |
| AE-NL | .5951 | .5329 | .7207 | .7027 | .5050 | .7173 | .5328 | .7011 | .5454 | .5031 | .8323* | .8486* | .7196 | .7196 | .8212 | .8270 | .8117 | .8143 | **.8331** | **.8513** | 0.10%↑ | 0.32%↑ |
| AE-US | .5086 | .5881 | .6114 | .6808 | .5395 | .6358 | .4627 | .5793 | .5387 | .5447 | .8385* | .8385* | .7625 | .7625 | .8334 | .8390 | .8433 | .8404 | **.8484** | **.8629** | 1.18%↑ | 2.91%↑ |

information privacy, this industrial dataset has been encrypted and desensitized, which is only used for academic research. For a clear comparison, we list the dataset statistics in Table II.

*2) Parameter Setting:* To ensure a fair comparison, we align the hyper-parameters of our DCMT with those baseline models according to the parameter settings reported in their original papers. For *Embedding Layer* of our DCMT model, we set the embedding dimension for each feature (sparse id feature or dense numerical feature or weighted feature) to 32 owing to the best performance observed from the experiment (see Fig. 8(a)). For the deep towers in *CTR Task* and *CVR Task* in Fig. 3, the MLP structure of the layers is [64-64-32] for the AE datasets, [320-200-80] for the Ali-CCP dataset, and [512-256-128] for the industrial dataset (see Fig. 8(b)). For training our DCMT model, we adopt Adam [22] to train the neural networks and set the maximum number of training epochs to 5. The learning rate is $0.001$, the weight of the counterfactual regularizer $\lambda_1$ is $0.001$, the regularization coefficient $\lambda_2$ is $0.0001$, and the batch size is $1,024$.

*3) Evaluation Metrics:* For the offline test, since the main task for this work is to predict post-click conversion rate, in the experiments, we report the experimental results of *CVR Task* and *CTCVR Task*. The area under the ROC curve (AUC) is adopted as performance metrics in the experiments. All experiments are repeated 5 times and averaged results are reported. Also, for the online A/B test, we mainly focus on the three online business metrics, i.e., PV-CTR (the click-through rate for each page view), PV-CVR (the conversion rate for

each page view), and Top-5 PV-CVR (PV-CVR for the top-5 services, a maximum of 5 services can be displayed on one screen). From the perspective of business, enhancing the click-through rate and conversion (the double click) rate for each page view from users can significantly improve the user experience and attract more new users.

*4) Comparison Methods:* As shown in Table III, we compare our DCMT model with seven baseline models in three groups, i.e., (1) parallel MTL baselines, (2) multi-gate MTL baselines, and (3) causal baselines. All seven baselines are representative and/or state-of-the-art approaches for each group. To make an ablation study, we implement other two simplified versions as the variants of the completed DCMT, i.e., DCMT_PD (only considering the propensity-based debiasing in the factual exposure space $\mathcal{D}$) and DCMT_CF (only adopting the counterfactual mechanism). For a clear comparison, in Table III, we list the detailed structures, task relationships, and main ideas of all the models implemented in the experiments.

### B. Performance Comparison and Analysis

*1) Offline test on the public datasets (**Results 1 and 2**):*
**Result 1: Offline performance comparison (for Q1).** Table IV shows the experimental results in term of AUC in different datasets for *CVR Task* and *CTCVR Task*, respectively. To answer **Q1**, we compare the performance of our DCMT with those of the seven baseline models. As indicated in Table IV, our DCMT outperforms the best-performing baselines by an average improvement of $1.07\%$ in term of CVR AUC
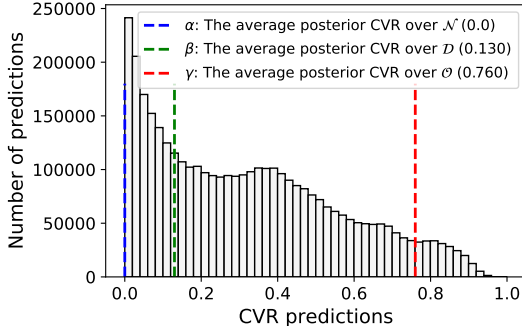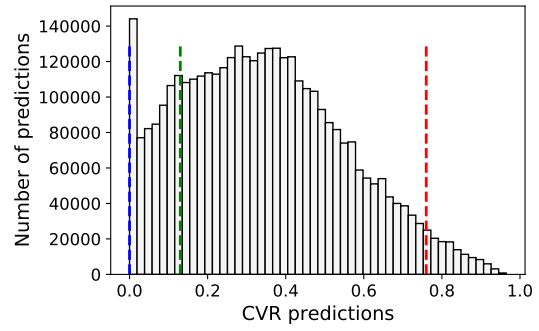
(a) The CVR prediction distribution of the base model (MMOE) over $\mathcal{D}$ (the average CVR prediction is 0.320 that is close to $\beta$)
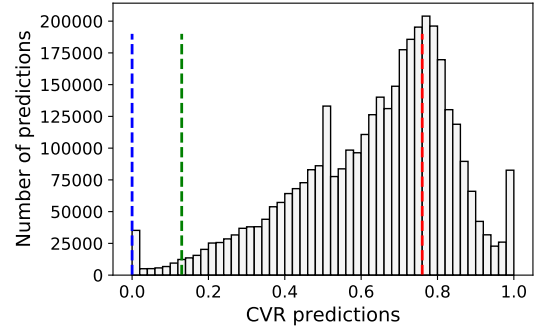
(b) The CVR prediction distribution of ESCM$^2$-IPW over $\mathcal{D}$ (the average CVR prediction is 0.676 that is close to $\gamma$)
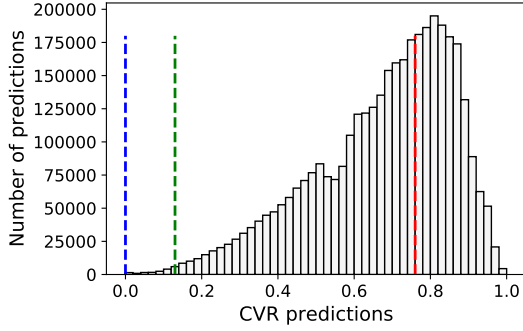
(c) The CVR prediction distribution of ESCM$^2$-DR over $\mathcal{D}$ (the average CVR prediction is 0.637 that is close to $\gamma$)

(d) The CVR prediction distribution of DCMT over $\mathcal{D}$ (the average CVR prediction is 0.343 that is close to $\beta$)

Fig. 7. The online prediction distribution of CVR over $\mathcal{D}$.

| Metric | Model | Day1 | Day2 | Day3 | Day4 | Day5 | Day6 | Day7 | Overall |
|---|---|---|---|---|---|---|---|---|---|
| PV-CTR | ESCM$^2$-IPW | 0.47%↑ | 1.00%↓ | 0.14%↑ | 0.95%↓ | 0.43%↓ | 0.23%↓ | 0.28%↑ | 0.32%↓ |
| | ESCM$^2$-DR | 0.23%↓ | 0.43%↓ | 0.13%↑ | 0.61%↓ | 0.06%↓ | 0.39%↓ | 0.03%↓ | 0.24%↓ |
| | DCMT | 0.32%↓ | 0.44%↓ | 0.91%↑ | 0.36%↑ | 0.99%↑ | 0.67%↑ | 1.26%↑ | 0.49%↑ |
| PV-CVR | ESCM$^2$-IPW | 0.43%↑ | 1.15%↓ | 0.25%↑ | 0.91%↓ | 0.42%↓ | 0.00% | 0.68%↑ | 0.22%↓ |
| | ESCM$^2$-DR | 0.04%↓ | 0.28%↓ | 0.44%↑ | 0.57%↓ | 0.26%↑ | 0.31%↓ | 0.21%↓ | 0.02%↓ |
| | DCMT | 0.51%↑ | 0.42%↓ | 1.49%↑ | 0.53%↑ | 0.89%↑ | 0.71%↑ | 1.42%↑ | 0.75%↑ |
| Top-5 PV-CVR | ESCM$^2$-IPW | 0.38%↑ | 1.09%↓ | 0.26%↑ | 0.85%↓ | 0.41%↓ | 0.01%↑ | 0.66%↑ | 0.21%↓ |
| | ESCM$^2$-DR | 0.07%↓ | 0.29%↓ | 0.47%↑ | 0.48%↓ | 0.26%↑ | 0.35%↓ | 0.20%↓ | 0.02%↓ |
| | DCMT | 0.47%↑ | 0.67%↓ | 1.59%↑ | 0.58%↑ | 0.88%↑ | 0.74%↑ | 1.51%↑ | 0.66%↑ |

(the main task). In particular, our DCMT improves the best-performing baselines (with results marked by * in Table IV) by 2.56% on the Ali-CCP dataset, 0.88% on the AE-ES dataset, 0.61% on AE-FR, 0.10% on AE-NL, and 1.18% on AE-US. This is because our DCMT can debias the selection bias over $\mathcal{D}$, demonstrating the superiority in terms of easing selection bias and data sparsity.

**Result 2: Ablation study (for Q2).** To answer **Q2**, we implement two variants of our DCMT, i.e., DCMT_PD (only considering the propensity-based debiasing in $\mathcal{D}$) and DCMT_CF (only adopting the counterfactual mechanism) to demonstrate

the detailed contributions of the two main components of our DCMT framework.

On the one hand, as we can see from Table IV, when we only consider the propensity-based debiasing (DCMT_PD) in $\mathcal{D}$, our completed DCMT improves DCMT_PD by an average of 2.89% for CVR AUC. This result indicates that our counterfactual mechanism can further improve the prediction performance and effectively alleviate the problems of selection bias and data sparsity.

On the other hand, compared with DCMT_CF, our completed DCMT model achieves an average improvement of 1.91% (according to the results in Table IV). This means that the propensity-based debiasing (DCMT_PD) in $\mathcal{D}$ also plays a very important role in our DCMT.

*2) Online A/B test on the Alipay platform (**Result 3**):*
**Result 3-1: Online performance comparison (for Q3).** To answer **Q3**, in addition to the above-mentioned offline experiments, we also conduct extensive online A/B experiments on the Alipay Search platform to further validate the performance of our DCMT framework. In the online setting, an MMOE-based model is the base model (the current running model), and we also choose two state-of-the-art models, i.e., ESCM$^2$-IPW, and ESCM$^2$-DR, as the online baselines. Specifically, we first implement these baselines and our DCMT with our TensorFlow-based machine learning framework and deploy them on our inference platform. We then randomly assign the same number of users into the corresponding four buckets,
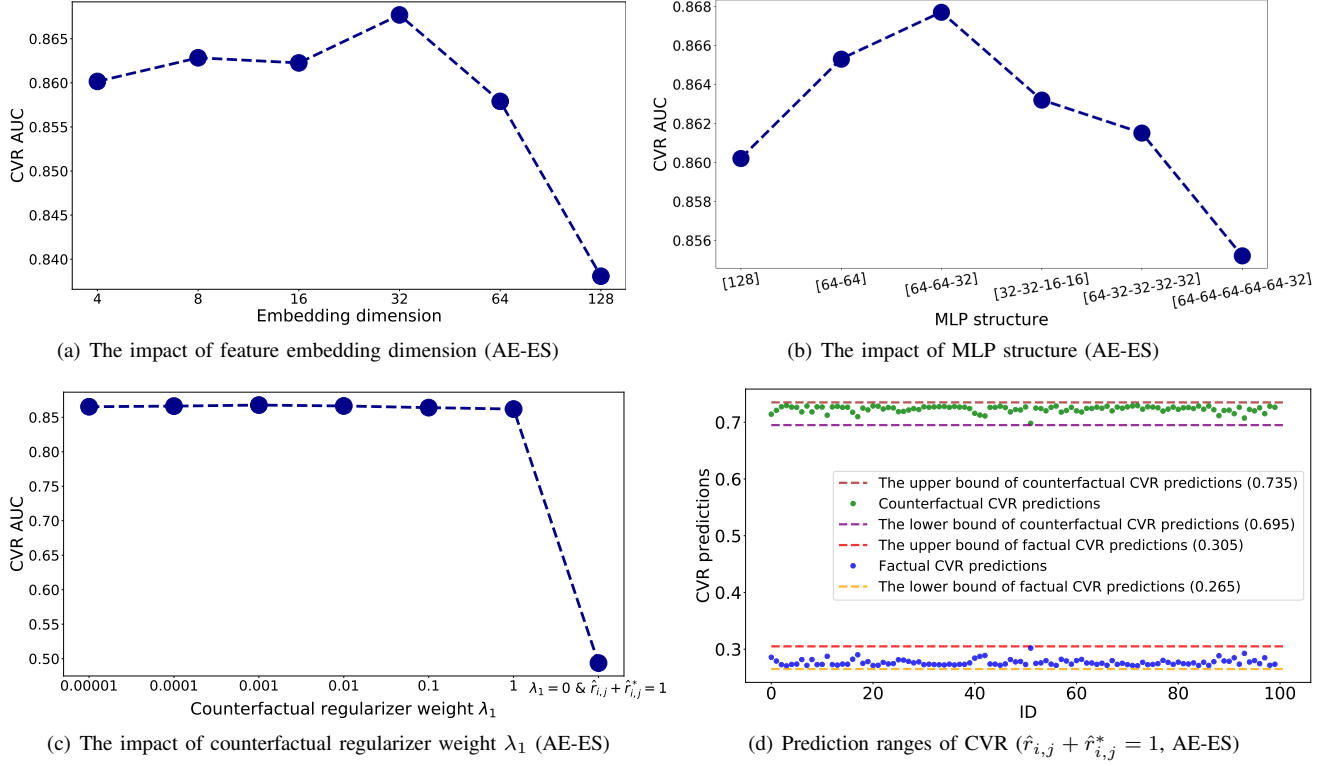
(a) The impact of feature embedding dimension (AE-ES)



(b) The impact of MLP structure (AE-ES)



(c) The impact of counterfactual regularizer weight $\lambda_1$ (AE-ES)



(d) Prediction ranges of CVR ($\hat{r}_{i,j} + \hat{r}_{i,j}^* = 1$, AE-ES)

Fig. 8. Impact of Hyper-parameters.

i.e., MMOE (the base model), ESCM$^2$-IPW, ESCM$^2$-DR, and DCMT, via using our online A/B testing platform. Finally, we observe each model's online performance in the respective bucket of users on the A/B testing platform. From the perspective of business, we mainly focus on the three online metrics, i.e., PV-CTR, PV-CVR, and Top-5 PV-CVR. Finally, we report the online A/B testing results of a week (in August 2022) in Table V.

This online experiment lasts 7 days and covers around 4.1 million unique visitors (UVs) and around 8.9 million page views (PVs) for each experiment bucket. Overall, our DCMT framework improves the base model (MMOE) by 0.49% (for PV-CTR) with 95% confidence intervals, 0.75% (for PV-CVR) with 95% confidence intervals, and 0.66% (for Top-5 PV-CVR) with 95% confidence intervals. Also, as we can see from Table V, our DCMT can consistently outperform both ESCM$^2$-IPW and ESCM$^2$-DR in terms of PV-CTR, PV-CVR, and Top-5 PV-CVR.

**Result 3-2: Online prediction comparison (for Q3).** To clearly compare the CVR prediction performance of our DCMT with those of the two online baselines, i.e., ESCM$^2$-IPW, and ESCM$^2$-DR, we collect the online predictions of the four online methods over the infer space $\mathcal{D}$ from the log of Day 1 on our online A/B testing platform and draw their distributions in Figure 7. Note that in the online environment, we cannot obtain the ground-truth distribution of CVR predictions. Thus, to compare the CVR prediction performance of the four online methods, we mark the average

posterior CVR over $\mathcal{D}$, $\mathcal{O}$, and $\mathcal{N}$ respectively in Figure 7. As we can see from Figure 7, the average CVR predictions of ESCM$^2$-IPW (0.676), and ESCM$^2$-DR (0.637) over $\mathcal{D}$ are close to the average posterior CVR over $\mathcal{O}$ (0.760) and away from the average posterior CVR over $\mathcal{D}$ (0.130). This means that only debiasing over $\mathcal{O}$ (ESCM$^2$-IPW and ESCM$^2$-DR) cannot infer very well over $\mathcal{D}$. Although the average CVR prediction of the base model (MMOE) is close to the average posterior CVR over $\mathcal{D}$ (0.130), there are many predictions concentrating between the average posterior CVR over $\mathcal{N}$ (0.0) and the average posterior CVR over $\mathcal{N}$ (0.130). In contrast, the prediction majority of our DCMT concentrates between the average posterior CVR over $\mathcal{D}$ (0.130) and the average posterior CVR over $\mathcal{O}$ (0.760), which is significantly better than those of the other three online baselines. This result demonstrates that our DCMT can effectively debias the selection bias in $\mathcal{D}$.

### C. Impact of Hyper-parameters (Result 4)

To answer **Q4**, in this section, we analyse the impacts of the dimension of feature embedding, the structure of MLP in our DCMT, and the weight of the counterfactual regularizer $\lambda_1$. Due to space limitations, we only report the offline experimental results of the CVR task on the AE-ES dataset.

**Impact of feature embedding dimension.** To study the impact of feature embedding dimension on our DCMT framework, we choose different dimensions, i.e., $\{4, 8, 16, 32, 64, 128\}$, for performance comparison. The results of *CVR Task* on AE-ES are reported in Fig. 8(a). As

observed from Fig. 8(a), when the embedding dimension is 16, our DCMT model achieves the best performance. From 16 to 128, the performance decreases with the embedding dimension because a large dimension may make our DCMT framework over-fitting to the training samples.

**Impact of MLP structure.** To study the impact of MLP structure on our DCMT framework, we attempt to set the depth of MLP in the deep towers (see Fig. 3) from 1 to 6 with different numbers of units. Due to space limitations, we only report the best-performing structure for each depth, e.g., $[128]$ for $d_{mlp} = 1$ and $[64 - 64]$ for $d_{mlp} = 2$. Similar to the above experiment of embedding dimension, we only report the AUC results of *CVR Task* on AE-ES in Fig. 8(b). As we can see from Fig. 8(b), when the structure of MLP is $[64 - 64 - 32]$ ($d_{mlp} = 3$), our DCMT can achieve the best performance. In general, the performance of our DCMT increases with the depth of MLP from 1 to 3. After that, the performance will decrease mainly because a complex MLP structure may lead to an over-fitting problem. Similarly, we also choose the best-performing structures for the Ali-CCP dataset ($[320 - 200 - 80]$) and the Alipay Search dataset ($[512 - 256 - 128]$).

**Impact of counterfactual regularizer weight $\lambda_1$.** To study the impact of counterfactual regularizer weight $\lambda_1$ in our DCMT framework, we choose different weights, i.e., $\{0.00001, 0.0001, 0.001, 0.01, 0.1, 1\}$, for the counterfactual regularizer $L$ in Eq. (9). In addition to the soft constraint, we also conduct an experiment to validate the performance of the hard constraint, i.e., we force $\hat{r}_{i,j} + \hat{r}_{i,j}^* = 1$. Similarly, we only report the experimental results on the AE-ES dataset in Fig. 8(c). In general, the performance of our DCMT increases from 0.00001 to 0.001 (the best-performing weight). After that, the performance will decrease mainly because a stronger constraint may make the training process of our CVR estimation difficult to minimize the main loss, i.e., the loss between CVR labels and CVR predictions. Also, the CVR AUC of the hard constraint, i.e., $\lambda_1 = 0$ and $\hat{r}_{i,j} + \hat{r}_{i,j}^* = 1$, is significantly worse than those of the soft constraints in Fig. 8(c). To study this issue, we randomly choose 100 samples in the AE-ES dataset, and report their factual CVR predictions and counterfactual CVR predictions in Fig. 8(d). The factual CVR predictions and counterfactual CVR predictions of these samples are restricted in small value ranges, i.e., $[0.265, 0.305]$ and $[0.695, 0.735]$. This issue prevents our DCMT to minimize the main loss and thus we did not choose the hard constraint, which has been analysed in Section III-C as well.

## V. Related Work

### A. Multi-Task Learning for CVR Prediction

Click-through rate (CTR) and conversion rate (CVR) predictions are two traditional tasks in recommender systems, and some of existing approaches [1], [7], [10], [23]–[29] in the literature attempt to understand users' click and purchase behaviour. In [1], Ma et al. firstly proposed click-through & conversion rate (CTCVR) as an auxiliary task with CTR task to

help improve the prediction accuracy of CVR. Since the scope of this paper is post-click CVR prediction, we only review the highly-related literature on CVR prediction and multiple tasks (including CVR prediction) as follows.

**CVR Prediction.** CVR prediction is a very challenging task because conversed (purchased) samples are very rare in real-world datasets. Recently, with help of deep neural network, there are some related solutions [1], [2], [9], [10], [23], [29]–[33] focusing on CVR prediction. These approaches adopt some effective feature representation strategies and end-to-end models to study the underlying logic between conversion and input features, e.g., user profiles, item details, exposure position, and users' behaviour.

**Multiple Tasks.** In [1], to address selection bias and data sparsity problems, the proposed model, i.e., Entire Space Multi-task Model (ESMM), is trained in the entire exposure space. Inevitably, the ESMM may suffer from other bias problems, e.g., exposure bias. The extension works [2], [9], [10], [23], [29] further improve the prediction accuracies of CVR and CTCVR by considering more missing samples and users' feedback actions. Also, some traditional and novel multi-task learning framework, e.g., Cross Stitch [12], MMOE [13], PLE [14], and MOSE [15], are adopted to predict CTR, CVR, and CTCVR, simultaneously.

### B. Causal Inference for Recommendation

Causal inference is to study the independent and actual effect of a particular component in a system [34]. Causal inference has been utilized in some recommender systems, aiming to accurately infer users' preferences for items. Early studies [35]–[37] mainly focus on debiasing implicit feedback, e.g., position bias [38]. Next, some researchers start to study the recommendation fairness [39] and other biases, e.g., exposure bias [40], popularity bias [41], in the training samples. Also, there is a novel work, i.e., counterfactual recommendation [42]–[44], which adopts counterfactual inference to mitigate the clickbait problem in recommender systems. To further debias the selection bias in the training samples, Inverse Propensity Scoring Weighting (IPW) [45] is applied to regenerate the original samples based on a re-weighting strategy. With the help of multi-task learning, recently, some IPW-based debiasing approaches, e.g., Multi-IPW [10] and ESCM$^2$-IPW [5], are proposed for CVR estimation.

## VI. Conclusion and Future Work

In this paper, we have proposed a **D**irect entire-space **C**ausal **M**ulti-**T**ask framework, namely DCMT, for post-click conversion prediction. To debias the selection bias in the entire exposure space $\mathcal{D}$, we propose a new counterfactual mechanism with our proposed structure of twin tower, which can effectively improve the prediction accuracy and avoid the over-fitting problem caused by selection bias and data sparsity. Also, we have conducted extensive experiments to demonstrate the superior performance of our proposed DCMT model. In the future, we plan to study the effect of different counterfactual strategies on our DCMT's performance.

REFERENCES

[1] X. Ma, L. Zhao, G. Huang, Z. Wang, Z. Hu, X. Zhu, and K. Gai, "Entire space multi-task model: An effective approach for estimating post-click conversion rate," in *SIGIR*, 2018, pp. 1137–1140.

[2] H. Wen, J. Zhang, Y. Wang, F. Lv, W. Bao, Q. Lin, and K. Yang, "Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction," in *SIGIR*, 2020, pp. 2377–2386.

[3] Y. Zhang, H. Dai, C. Xu, J. Feng, T. Wang, J. Bian, B. Wang, and T.-Y. Liu, "Sequential click prediction for sponsored search with recurrent neural networks," in *AAAI*, vol. 28, no. 1, 2014.

[4] M. Richardson, E. Dominowska, and R. Ragno, "Predicting clicks: estimating the click-through rate for new ads," in *WWW*, 2007, pp. 521–530.

[5] H. Wang, T.-W. Chang, T. Liu, J. Huang, Z. Chen, C. Yu, R. Li, and W. Chu, "Escm$^2$: Entire space counterfactual multi-task model for post-click conversion rate estimation," *arXiv preprint arXiv:2204.05125*, 2022.

[6] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu, and K. Gai, "Deep interest evolution network for click-through rate prediction," in *AAAI*, vol. 33, no. 01, 2019, pp. 5941–5948.

[7] P. Li, Z. Jiang, M. Que, Y. Hu, and A. Tuzhilin, "Dual attentive sequential learning for cross-domain click-through rate prediction," in *SIGKDD*, 2021, pp. 3172–3180.

[8] J. Pan, J. Xu, A. L. Ruiz, W. Zhao, S. Pan, Y. Sun, and Q. Lu, "Field-weighted factorization machines for click-through rate prediction in display advertising," in *WWW*, 2018, pp. 1349–1357.

[9] D. Xi, Z. Chen, P. Yan, Y. Zhang, Y. Zhu, F. Zhuang, and Y. Chen, "Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising," *arXiv preprint arXiv:2105.08489*, 2021.

[10] W. Zhang, W. Bao, X.-Y. Liu, K. Yang, Q. Lin, H. Wen, and R. Ramezani, "Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning," in *WWW*, 2020, pp. 2775–2781.

[11] P. Wu, H. Li, Y. Deng, W. Hu, Q. Dai, Z. Dong, J. Sun, R. Zhang, and X.-H. Zhou, "On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges." IJCAI, 2022.

[12] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *CVPR*, 2016, pp. 3994–4003.

[13] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *SIGKDD*, 2018, pp. 1930–1939.

[14] H. Tang, J. Liu, M. Zhao, and X. Gong, "Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations," in *RecSys*, 2020, pp. 269–278.

[15] Z. Qin, Y. Cheng, Z. Zhao, Z. Chen, D. Metzler, and J. Qin, "Multitask mixture of sequential experts for user activity streams," in *SIGKDD*, 2020, pp. 3083–3091.

[16] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir *et al.*, "Wide & deep learning for recommender systems," in *DLRS*, 2016, pp. 7–10.

[17] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and antimatter in the universe," *New Journal of Physics*, vol. 14, no. 9, p. 095012, 2012.

[18] S. L. Morgan and C. Winship, *Counterfactuals and causal inference*. Cambridge University Press, 2015.

[19] A. Swaminathan and T. Joachims, "The self-normalized estimator for counterfactual learning," *advances in neural information processing systems*, vol. 28, 2015.

[20] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims, "Recommendations as treatments: Debiasing learning and evaluation," in *international conference on machine learning*. PMLR, 2016, pp. 1670–1679.

[21] pengcheng Li, R. Li, Q. Da, A.-X. Zeng, and L. Zhang, "Improving multi-scenario learning to rank in e-commerce by exploiting task relationships in the label space," in *CIKM*. New York,NY,USA: ACM, 2020.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] J. Pan, Y. Mao, A. L. Ruiz, Y. Sun, and A. Flores, "Predicting different types of conversions with multi-task learning in online advertising," in *SIGKDD*, 2019, pp. 2689–2697.

[24] W. Ouyang, X. Zhang, L. Li, H. Zou, X. Xing, Z. Liu, and Y. Du, "Deep spatio-temporal neural networks for click-through rate prediction," in *SIGKDD*, 2019, pp. 2078–2086.

[25] B. Liu, C. Zhu, G. Li, W. Zhang, J. Lai, R. Tang, X. He, Z. Li, and Y. Yu, "Autofis: Automatic feature interaction selection in factorization models for click-through rate prediction," in *SIGKDD*, 2020, pp. 2636–2645.

[26] Q. Song, D. Cheng, H. Zhou, J. Yang, Y. Tian, and X. Hu, "Towards automated neural interaction discovery for click-through rate prediction," in *SIGKDD*, 2020, pp. 945–955.

[27] Z. Meng, J. Zhang, Y. Li, J. Li, T. Zhu, and L. Sun, "A general method for automatic discovery of powerful interactions in click-through rate prediction," *arXiv preprint arXiv:2105.10484*, 2021.

[28] F. Zhu, Y. Wang, C. Chen, J. Zhou, L. Li, and G. Liu, "Cross-domain recommendation: challenges, progress, and prospects," *arXiv preprint arXiv:2103.01696*, 2021.

[29] H. Wen, J. Zhang, F. Lv, W. Bao, T. Wang, and Z. Chen, "Hierarchically modeling micro and macro behaviors via multi-task learning for conversion rate prediction," *arXiv preprint arXiv:2104.09713*, 2021.

[30] Q. Lu, S. Pan, L. Wang, J. Pan, F. Wan, and H. Yang, "A practical framework of conversion rate prediction for online display advertising," in *ADKDD*, 2017, pp. 1–9.

[31] W. Bao, H. Wen, S. Li, X.-Y. Liu, Q. Lin, and K. Yang, "Gmcm: Graph-based micro-behavior conversion model for post-click conversion rate estimation," in *SIGIR*, 2020, pp. 2201–2210.

[32] Y. Su, L. Zhang, Q. Dai, B. Zhang, J. Yan, D. Wang, Y. Bao, S. Xu, Y. He, and W. Yan, "An attention-based model for conversion rate prediction with delayed feedback via post-click calibration." in *IJCAI*, 2020, pp. 3522–3528.

[33] H. Wen, J. Zhang, Q. Lin, K. Yang, and P. Huang, "Multi-level deep cascade trees for conversion rate prediction in recommendation system," in *AAAI*, vol. 33, no. 01, 2019, pp. 338–345.

[34] J. Pearl, "Causal inference in statistics: An overview," *Statistics surveys*, vol. 3, pp. 96–146, 2009.

[35] D. Liang, L. Charlin, and D. M. Blei, "Causal inference for recommendation," in *AUAI*, 2016.

[36] T. Joachims, A. Swaminathan, and T. Schnabel, "Unbiased learning-to-rank with biased feedback," in *WSDM*, 2017, pp. 781–789.

[37] Q. Ai, K. Bi, C. Luo, J. Guo, and W. B. Croft, "Unbiased learning to rank with unbiased propensity estimation," in *SIGIR*, 2018, pp. 385–394.

[38] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey, "An experimental comparison of click position-bias models," in *WSDM*, 2008, pp. 87–94.

[39] M. Morik, A. Singh, J. Hong, and T. Joachims, "Controlling fairness and bias in dynamic learning-to-rank," in *SIGIR*, 2020, pp. 429–438.

[40] D. Liang, L. Charlin, J. McInerney, and D. M. Blei, "Modeling user exposure in recommendation," in *WWW*, 2016, pp. 951–961.

[41] H. Abdollahpouri, R. Burke, and B. Mobasher, "Managing popularity bias in recommender systems with personalized re-ranking," in *FLAIRS*, 2019.

[42] W. Wang, F. Feng, X. He, H. Zhang, and T.-S. Chua, "Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue," in *SIGIR*, 2021, pp. 1288–1297.

[43] J.-P. Fang, J. Zhou, Q. Cui, C.-Z. Tang, and L.-F. Li, "Interpreting model predictions with constrained perturbation and counterfactual instances," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 36, no. 01, p. 2251001, 2022.

[44] C. Tang, H. Wang, X. Li, Q. Cui, Y.-L. Zhang, F. Zhu, L. Li, J. Zhou, and L. Jiang, "Debiased causal tree: Heterogeneous treatment effects estimation with unmeasured confounding," in *Advances in Neural Information Processing Systems*, 2022.

[45] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.