Joint Training of Propensity Model and Prediction Model via Targeted Learning for Recommendation on Data Missing Not at Random

Hao Wang[1]

[1] Carnegie Mellon University
haow2@alumni.cmu.edu

Abstract

Recommender systems (RS) help to capture users' personalized interests and are increasingly important across social media, e-commerce, and various online applications. Since users are free to choose items to rate, the collected ratings are a missing-not-at-random subset of all user-item pairs, and there is a systematic distributional shift between observed and unseen ratings, which is also called the selection bias. There have been emerging quantities of methods to address the selection bias. The error-imputation-based, inverse propensity score, and doubly robustness methods try to improve the prediction accuracy by introducing the imputation model and propensity model. However, most of these methods cannot achieve nonparametric efficiency in estimating the ideal loss or lack theoretical guarantees for robustness and efficiency when learning the prediction model. To bridge this gap, this paper uses a neural network based architecture to model the propensity and prediction model and jointly train the two models with a target learning approach. Specifically, we add a targeted regularization that guides the optimization in the most efficient direction. Experiments on three widely used real-world datasets show the effectiveness of our method.

Introduction

Recommender systems (RS) is an effective tool to address information overload in the modern era, which has become increasingly important and widely applied in social e-commerce, entertainment, social media, and other online applications (Shi, Larson, and Hanjalic 2014; Zhang et al. 2023; Wang et al. 2020a). The RS aims to combine the features of users and items with the collected historical behaviors or feedback to provide personalized recommendations. However, as users are free to choose items to rate and the collected data is easily affected by item popularity differences and user rating preferences, the observed ratings or interactions have an inherent distributional shift from the target population, i.e., the ratings for all user-item pairs (Marlin et al. 2007; Marlin and Zemel 2009; Wang et al. 2024; Yang et al. 2023; Xiao et al. 2024; Wang et al. 2021). The bias due to the missing-not-at-random (MNAR) ratings is called selection bias. Training a prediction model from these MNAR ratings using empirical risk minimization (ERM) cannot achieve optimal prediction performance on all user-item pairs.

Many debiasing methods have been proposed to address selection bias. The error-imputation-based (EIB) methods tackle the selection bias by using an imputation model to estimate the missing labels and then minimizing the error between the observed and imputed pseudo-labels (Steck 2010). However, the performance of EIB highly relies on the accurate imputation. Instead, the inverse propensity score (IPS) methods address the selection bias by using a reweighting strategy to adjust the distribution of observed data to that of the missing data (Saito et al. 2020; Schnabel et al. 2016; Wang et al. 2022). Though the IPS method does not rely on an extrapolation model like EIB, due to the data sparsity, it is difficult to accurately estimate the propensity scores and may suffer from large variance when there exist small propensity scores. The doubly robust (DR) methods utilize both the imputation model and the propensity models (Saito 2020; Wang et al. 2019). It possesses advantages for (1) obtaining unbiased estimations when either the imputation model or the propensity model is correctly specified, which is known as the double robustness property, and (2) possessing lower variance than IPS. However, though doubly robust, the DR methods may still suffer from low efficiency and have large variances inherited from IPS modules due to the data sparsity. Furthermore, the double robustness of DR only holds under the strict correctness of either model, in other words, if the propensity model is slightly incorrect, the DR methods will suffer from a large bias (Kang and Schafer 2007; Molenberghs et al. 2014; Vermeulen and Vansteelandt 2015; Seaman and Vansteelandt 2018).

To address the above concerns and improve efficiency and robustness, we proposed a targeted learning method to reduce the asymptotic variance and guarantee robustness under model misspecification. Targeted learning aims to train the model that can meet different requirements like debiasing, variance control, or covariate balancing (Van der Laan, Rose et al. 2011), which is a flexible framework that has wide applications. The EIB, IPS, and DR methods have multiple models, like the prediction models, the propensity models, and the imputation models. Thus, it is hard to correctly specify all models. Meanwhile, it is important to reduce the variance and guarantee robustness simultaneously.

In this paper, we propose a joint training approach for propensity and prediction models via targeted learning, called **TLNet**. We propose a two-headed neural network to simul-

taneously learn the propensity and prediction models. We share the feature representation module across two models, which effectively addresses the data sparsity issue. We then propose a targeted regularization to impose the empirical efficient influence curve to 0, which guarantees robustness and efficiency for the estimation of the prediction model based on the theory of targeted learning. Specifically, we jointly learn the propensity model and prediction model with a combined loss of the targeted regularization and the prediction loss. The main contributions of this paper are as below:

- We propose a target learning network to reduce the variance and guarantee robustness and estimation efficiency simultaneously.
- In this network, we learn the sharing representation between the propensity and prediction models and propose a targeted regularization to impose the empirical efficient influence curve to 0
- Extensive experiments are conducted on three public datasets, demonstrating the superiority of applying target learning for debiasing recommender systems.

Related Work

Causal Recommendation

The causal recommendation aims to formulate the recommendation as a causal problem and debias the recommendation from the causal perspective (Wang et al. 2023; Yang et al. 2021, 2023). Causal recommendation plays an important role in debiased recommendations due to its good theoretical properties. There have been various causal recommendation methods proposed for unbiased recommendation under different types of bias, such as popularity bias (Zhang et al. 2021), user self-selection bias (Saito 2020), position bias (Ai et al. 2018), and model selection bias (Yuan et al. 2019). Most of them are IPS-based or DR-based methods.

IPS methods address the bias by adjusting the distribution of observed data to the distribution of unseen data, eliminating the distribution shift that is the major cause of bias. Schnabel et al. (2016) formulated the recommendation as treatment and denoised the prediction models with IPS and self-normalized IPS (SNIPS). Saito et al. (2020) further extended the IPS method to implicit recommendations. However, IPS methods are easily affected by misspecified propensity models and have large variances due to data sparsity issues. The DR methods have better properties like double robustness and lower variance. A series of enhanced DR methods have been proposed, including Multi-DR (Zhang et al. 2020), MRDR (Guo et al. 2021), DR-MSE (Dai et al. 2022), BRD-DR (Ding et al. 2022), SDR (Li, Zheng, and Wu 2023), DR-V2 (Li et al. 2023b), CDR (Song et al. 2023), KBDR (Li et al. 2024d), N-DR (Li et al. 2024b), DCE-DR (Kweon and Yu 2024), DT-DR (Zhang et al. 2024), UIDR (Li et al. 2024c), and OME-DR (Li et al. 2024d). Most of the existing enhanced methods aim to achieve a better bias-variance trade-off and are less robust, while we proposed the TLNet to guarantee the efficiency and robustness with targeted learning.

Targeted Learning

Targeted learning is a flexible framework in causal inference that can apply to multiple fields and lead to many field-specific approaches to address scientific problems in different fields, including survival analysis, genomics, and epidemiology. Targeted learning serves as a general framework and is model agnostic. Shi, Blei, and Veitch (2019) proposed Dragonnet to adapt the neural networks to estimate the treatment effects based on targeted learning and proved that the targeted learning could be well combined with the neural networks and achieve satisfying performance. TDR was the first method that extended the target learning to the field of debiased recommendation (Li et al. 2023a). TDR debiased the prediction model by dealing with the estimator and learning problem simultaneously. However, the TDR only considered applying targeted learning on the imputation model, which cannot directly guarantee efficiently estimating the prediction model that is the ultimate goal. The TLNet extended a more fine-grained targeted learning neural network to recommender systems and theoretically guarantees efficient estimation for the ideal loss and robust learning for the prediction model.

Problem Setup

Let $\mathcal{U} = \{u_1, \cdots, u_m\}$ be the users set, $\mathcal{I} = \{i_1, \cdots, i_n\}$ be the item set, and $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of all user-item pairs. The rating matrix is denoted as $\mathbf{R} \in \mathbb{R}^{m \times n}$ with $r_{u,i}$ as element. Let $o_{u,i} \in \{0, 1\}$ be the observation indicator indicating whether the $r_{u,i}$ is observed and $x_{u,i}$ be the feature. We denote the prediction model as $f_\theta(\cdot)$ parameterized by θ and the predicted ratings as $\hat{r}_{u,i} = f_\theta(x_{u,i})$. The goal is to accurately predict $r_{u,i}$ for all user-item pairs, which can be achieved by minimizing the ideal loss

$$\mathcal{L}_{\text{ideal}}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \mathcal{L}(f_\theta(x_{u,i}), r_{u,i}) := \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} e_{u,i},$$

where $\mathcal{L}(\cdot, \cdot)$ is the training loss function such as cross-entropy loss. However, in practice, we cannot obtain the complete rating matrix. We denote the set of user-item pairs with observed ratings as $\mathcal{O} = \{(u, i) \mid o_{u,i} = 1\}$. The naive method optimizes the average loss over the observed samples

$$\mathcal{L}_{\text{N}}(\theta) = \frac{1}{|\mathcal{O}|} \sum_{(u,i)\in\mathcal{O}} e_{u,i}.$$

Due to the selection bias, $\mathbb{E}\{\mathcal{L}_{\text{N}}(\theta)\} \neq \mathcal{L}_{\text{ideal}}(\theta)$ (Schnabel et al. 2016; Wang et al. 2019). Several methods were proposed to unbiasedly estimate the ideal loss, including the EIB, IPS, DR, and their variants. Because EIB and IPS can be regarded as a special case of DR, we only introduce the DR methods here. The loss function of the vanilla DR method is formulated as

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

where $\hat{p}_{u,i}$ is the estimated propensity score for the true exposure probability $p_{u,i} = \Pr(o_{u,i} = 1 \mid x_{u,i})$, and $\hat{e}_{u,i}$ is the error for the imputation model $m(x_{u,i}; \phi)$, i.e., $\hat{e}_{u,i} = \mathcal{L}(m(x_{u,i}; \phi), \hat{r}_{u,i})$.

Proposed Method

We propose a joint training method for the propensity and prediction model via targeted learning, called TLNet, to enhance the debiasing performance on data MNAR. Different from Dragonnet, TLNet is adapted to recommender systems where there are no observations for negatively treated samples. The TLNet can simultaneously guarantee efficiency with low variance and robustness under the misspecification of models using targeted learning.

First, we design a two-headed neural network to learn the prediction model and propensity model simultaneously. Starting from $x_{u,i}$, we use a shared feature representation module, i.e., $z_{u,i} = \Phi(x_{u,i})$. The hope is that the representation module can distill the covariates into the features relevant to the ratings and observation indicator. We then model the prediction model and propensity model with two different heads. Following previous notations, we denote the prediction model by $f_\theta(x_{u,i})$, in which θ includes the parameters of both the representation module and head for prediction model, and we denote the propensity model by $g_\vartheta(x_{u,i})$, in which ϑ includes the parameters of both the representation module and head for propensity model.

We expect that the prediction model can estimate the ratings well and the propensity model can estimate the observation indicators well. Thus, a straightforward loss function can be derived as

$$\mathcal{R}(\theta, \vartheta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{O}} \frac{\mathcal{L}(f_\theta(x_{u,i}), r_{u,i})}{g_\vartheta(x_{u,i})} + \alpha \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \tilde{\mathcal{L}}(g_\vartheta(x_{u,i}), o_{u,i}),$$

where $\alpha \in \mathbb{R}^+$ is a hyperparameter weighting the loss components, and $\mathcal{L}$ and $\tilde{\mathcal{L}}$ is the training loss function (such as cross-entropy loss) for prediction and propensity model, respectively. However, the above combined loss cannot guarantee the efficiency of the estimation for the prediction model.

We turn to find a target to control the efficiency of the estimation for the ideal loss. An intuitive way is to find the efficient influence curve for the ideal loss function and impose it as 0. We denote the expectation of the ideal loss by ψ, i.e.,

$$\psi = \mathbb{E}_\mathcal{D} \left[\mathcal{L}(f_\theta(x_{u,i}, r_{u,i})) \right],$$

and $\mathcal{L}_{\text{ideal}}(\theta)$ is equivalent to ψ under finite samples. The efficient influence curve for ψ can be derived as

$$\phi(x_{u,i}, r_{u,i}; \theta, \vartheta, \psi) = f_\theta(x_{u,i}) + \frac{o_{u,i}}{g_\vartheta(x_{u,i})} \times \{r_{u,i} - f_\theta(x_{u,i})\} - \psi.$$

Thus, the target is to impose the sum of $\phi(x_{u,i}, r_{u,i}; \theta, \vartheta, \psi)$ to be 0, that is,

$$0 = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \phi(x_{u,i}, r_{u,i}; \theta, \vartheta, \psi). \tag{1}$$

However, the optimization of the above target function is intractable. Thus, the optimization is modified to an equivalent form that is easier to optimize following Shi, Blei, and Veitch (2019). We first introduce an auxiliary learnable parameter ε and a regularization term $\gamma_{u,i}$ defined by

$$\tilde{G}(x_{u,i}, o_{u,i}; \theta, \vartheta, \varepsilon) = f_\theta(x_{u,i}) + \varepsilon \left[\frac{o_{u,i}}{g_\vartheta(x_{u,i})} \right],$$
$$\gamma(x_{u,i}, o_{u,i}, r_{u,i}; \theta, \vartheta, \varepsilon) = (r_{u,i} - \tilde{G}_{u,i})^2.$$

The target is to let $\frac{1}{|\mathcal{O}|} \sum_{(u,i)\in\mathcal{O}} \gamma(x_{u,i}, o_{u,i}, r_{u,i}; \theta, \vartheta, \varepsilon) = 0$. We train it as a regularization. The final loss is

$$\mathcal{L}_{\text{combined}}(\theta, \vartheta, \varepsilon) = \mathcal{R}(\theta, \vartheta) + \beta \frac{1}{|\mathcal{O}|} \sum_{(u,i)\in\mathcal{O}} \gamma(x_{u,i}, o_{u,i}, r_{u,i}; \theta, \vartheta, \varepsilon),$$

where β is a hyperparameter as the penalty coefficient. We learn parameters by minimizing the combined loss,

$$\hat{\theta}, \hat{\vartheta}, \hat{\varepsilon} = \arg\min_{\theta, \vartheta, \varepsilon} \mathcal{L}_{\text{combined}}(\theta, \vartheta, \varepsilon).$$

We define the following estimators,

$$\hat{\psi} = \frac{1}{|\mathcal{O}|} \sum_{(u,i)\in\mathcal{O}} (r_{u,i} - \hat{G}(x_{u,i}, o_{u,i}))^2, \text{ where}$$
$$\hat{G}(x_{u,i}, o_{u,i}) = \tilde{G}(x_{u,i}, o_{u,i}; \hat{\theta}, \hat{\vartheta}, \hat{\varepsilon}).$$

Then we observe that

$$\partial_\varepsilon \mathcal{L}_{\text{combined}}(\theta, \vartheta, \varepsilon) \mid_{\hat{\varepsilon}} = \beta \frac{1}{|\mathcal{O}|} \sum_{(u,i)\in\mathcal{O}} \psi(x_{u,i}, r_{u,i}; \hat{\theta}, \hat{\vartheta}, \hat{\psi}).$$

Thus, minimizing the combined loss would force the estimated parameters satisfy the target function Eq. (1), which is the nonparametric influence curve for the ideal loss. Based on the nonparametric theory, we can conclude that our estimation will achieve nonparametric efficiency. By imposing the target function as Eq. (1), even when both $f_\theta(x_{u,i})$ and $g_\vartheta(x_{u,i})$ are misspecified, the bias of the loss are enforced to be close to 0, thus, leading to the robustness of TLNet under misspecifications of models.

Experiments

Datasets. To evaluate the debiasing performance, we conducted experiments on three real-world datasets, **Coat** (Schnabel et al. 2016), **Yahoo! R3** (Schnabel et al. 2016), and **KuaiRec** (Gao et al. 2022), which are widely used in debiased RS because all of them include both biased data and unbiased data. **Coat** dataset consists of 6,960 biased ratings in the training set and 4,640 unbiased ratings in the test set from 290 users and 300 items. The **Yahoo! R3** dataset includes 311,704 biased ratings in the training set from 15,400 users and 1,000 items and 54,000 unbiased ratings from the first 5,400 users. Each rating in both datasets are five-scale. We binarize them by letting ratings less than three to 0 and 1 otherwise. Additionally, we use an industrial dataset **KuaiRec** with 4,676,570 records for video watching ratio of 1,411 users and 3,327 videos. We randomly split 100 videos for each user for unbiased evaluation. We binarize the records by letting values less than two be 0 and 1 otherwise.

Baselines. We use the neural collaborative filtering (NCF) method as base model, and we compare our method with

Table 1: Performance on AUC, Recall@K, and NDCG@K on the unbiased test set of Coat, Yahoo! R3, and KuaiRec. The best results are bolded and the best baseline is underlined. * means statistical significance with p-value < 0.05.

	COAT			YAHOO! R3			KUAIREC		
Method	AUC	R@5	N@5	AUC	R@5	N@5	AUC	R@50	N@50
NCF	$0.762_{\pm 0.011}$	$0.441_{\pm 0.010}$	$0.623_{\pm 0.011}$	$0.682_{\pm 0.001}$	$0.451_{\pm 0.002}$	$0.674_{\pm 0.002}$	$0.835_{\pm 0.001}$	$0.691_{\pm 0.007}$	$0.643_{\pm 0.004}$
CVIB	$0.759_{\pm 0.002}$	$0.455_{\pm 0.006}$	$0.636_{\pm 0.003}$	$0.696_{\pm 0.001}$	$0.452_{\pm 0.003}$	$0.683_{\pm 0.002}$	$0.792_{\pm 0.003}$	$0.641_{\pm 0.004}$	$0.584_{\pm 0.004}$
DIB	$0.747_{\pm 0.004}$	$0.455_{\pm 0.007}$	$0.643_{\pm 0.004}$	$0.697_{\pm 0.003}$	$0.436_{\pm 0.006}$	$0.671_{\pm 0.005}$	$0.800_{\pm 0.006}$	$0.643_{\pm 0.006}$	$0.576_{\pm 0.005}$
IPS	$0.758_{\pm 0.006}$	$0.448_{\pm 0.012}$	$0.636_{\pm 0.011}$	$0.690_{\pm 0.004}$	$0.452_{\pm 0.004}$	$0.674_{\pm 0.004}$	$0.836_{\pm 0.004}$	$0.692_{\pm 0.010}$	$0.647_{\pm 0.005}$
SNIPS	$0.759_{\pm 0.011}$	$0.449_{\pm 0.006}$	$0.644_{\pm 0.008}$	$0.692_{\pm 0.001}$	$0.450_{\pm 0.002}$	$0.677_{\pm 0.002}$	$0.834_{\pm 0.001}$	$\underline{0.703_{\pm 0.001}}$	$0.650_{\pm 0.001}$
AS-IPS	$0.759_{\pm 0.005}$	$0.446_{\pm 0.004}$	$0.633_{\pm 0.006}$	$0.688_{\pm 0.007}$	$0.454_{\pm 0.009}$	$0.674_{\pm 0.007}$	$0.836_{\pm 0.004}$	$0.694_{\pm 0.004}$	$0.644_{\pm 0.003}$
IPS-V2	$0.750_{\pm 0.007}$	$0.441_{\pm 0.004}$	$0.619_{\pm 0.005}$	$0.694_{\pm 0.007}$	$0.456_{\pm 0.003}$	$0.673_{\pm 0.005}$	$0.838_{\pm 0.005}$	$0.698_{\pm 0.006}$	$0.649_{\pm 0.005}$
DR-JL	$0.768_{\pm 0.011}$	$0.446_{\pm 0.005}$	$0.633_{\pm 0.007}$	$0.696_{\pm 0.003}$	$0.453_{\pm 0.003}$	$0.678_{\pm 0.002}$	$0.837_{\pm 0.001}$	$0.698_{\pm 0.003}$	$0.649_{\pm 0.003}$
MRDR-DL	$0.768_{\pm 0.009}$	$0.457_{\pm 0.012}$	$0.642_{\pm 0.013}$	$0.695_{\pm 0.002}$	$0.453_{\pm 0.003}$	$0.679_{\pm 0.003}$	$0.835_{\pm 0.002}$	$0.697_{\pm 0.005}$	$0.649_{\pm 0.003}$
DR-BIAS	$0.763_{\pm 0.010}$	$0.449_{\pm 0.005}$	$0.639_{\pm 0.007}$	$0.699_{\pm 0.001}$	$0.456_{\pm 0.003}$	$0.677_{\pm 0.003}$	$0.833_{\pm 0.004}$	$0.687_{\pm 0.007}$	$0.641_{\pm 0.002}$
DR-MSE	$0.764_{\pm 0.011}$	$0.452_{\pm 0.005}$	$0.635_{\pm 0.007}$	$0.698_{\pm 0.003}$	$\underline{0.456_{\pm 0.002}}$	$\underline{0.684_{\pm 0.002}}$	$0.833_{\pm 0.003}$	$0.692_{\pm 0.005}$	$0.643_{\pm 0.005}$
TDR-JL	$\underline{0.770_{\pm 0.008}}$	$\underline{0.455_{\pm 0.017}}$	$\underline{0.653_{\pm 0.024}}$	$\underline{0.701_{\pm 0.004}}$	$0.456_{\pm 0.005}$	$0.684_{\pm 0.006}$	$\underline{0.839_{\pm 0.002}}$	$0.696_{\pm 0.004}$	$0.651_{\pm 0.009}$
DR-V2	$0.761_{\pm 0.004}$	$0.452_{\pm 0.007}$	$0.635_{\pm 0.005}$	$0.690_{\pm 0.004}$	$0.451_{\pm 0.004}$	$0.682_{\pm 0.005}$	$0.836_{\pm 0.003}$	$0.700_{\pm 0.007}$	$0.649_{\pm 0.005}$
KBDR	$0.769_{\pm 0.008}$	$0.453_{\pm 0.005}$	$0.640_{\pm 0.006}$	$0.692_{\pm 0.007}$	$0.453_{\pm 0.008}$	$0.676_{\pm 0.004}$	$0.838_{\pm 0.006}$	$0.699_{\pm 0.008}$	$\underline{0.651_{\pm 0.004}}$
TLNet	$\mathbf{0.776}_{\pm 0.009}$	$\mathbf{0.477}^{*}_{\pm 0.014}$	$\mathbf{0.692}^{*}_{\pm 0.011}$	$\mathbf{0.702}_{\pm 0.003}$	$\mathbf{0.458}^{*}_{\pm 0.001}$	$\mathbf{0.687}^{*}_{\pm 0.002}$	$\mathbf{0.852}^{*}_{\pm 0.002}$	$\mathbf{0.716}^{*}_{\pm 0.003}$	$\mathbf{0.659}^{*}_{\pm 0.003}$

the following debiasing methods, including: (1) propensity-independent methods: **CVIB** (Wang et al. 2020b), **DIB** (Liu et al. 2021) and **AS-IPS** (Saito 2020); (2) IPS-based methods: **IPS** (Schnabel et al. 2016) and **SNIPS** (Schnabel et al. 2016), and **IPS-V2** (Li et al. 2023b); (3) DR-based methods: **DR-JL** (Wang et al. 2019), **MRDR-DL** (Guo et al. 2021), **DR-BIAS** (Dai et al. 2022), **DR-MSE** (Dai et al. 2022), **TDR-CL** (Li et al. 2023a), **DR-V2** (Li et al. 2023b), and **KBDR** (Li et al. 2024a).

Training Protocols and Details. We train all the methods on Pytorch with Adam as the optimizer. We tune the learning rate in $[0.005, 0.1]$, and λ in $[1e-6, 5e-3]$. We tune the trade-off parameter α and β in $\{0.1, 0.5, 1, 5, 10\}$. The batch size is chosen as 128 for **Coat** and 2048 for **Yahoo! R3** and **KuaiRec**. We evaluate the prediction performance with three widely adopted evaluation metrics: AUC, NDCG@K (N@K), and Recall@K (R@K). The N@K and R@K are popular in recommender systems as they can assess the quality of ranking tasks. Specifically, NDCG@K evaluates the quality of recommendations by considering the importance of each item's position based on discounted gains:

$$DCG_u@K = \sum_{i \in D^u_{\text{test}}} \frac{I(\hat{z}_{u,i} \leq K)}{\log(\hat{z}_{u,i}+1)},$$

$$NDCG@K = \frac{1}{|U|} \sum_{u \in U} \frac{DCG_u@K}{IDCG_u@K},$$

where IDCG represents the best possible DCG, D^u_{test} denotes the the cardinality of all ratings of the user u in test data, and $\hat{z}_{u,i}$ represents the ranking of item i in the recommended list for user u. In addition, the formula of Recall@K is as follows:

$$Recall_u@K = \frac{\sum_{i \in D^u_{\text{test}}} I(\hat{z}_{u,i} \leq k)}{\min(K, |D^u_{\text{test}}|)},$$

$$Recall@K = \frac{1}{|U|} \sum_{u \in U} Recall_u@K.$$

We set $K = 5$ for **Coat** and **Yahoo! R3**, and $K = 50$ for **KuaiRec**.

Experiment Results. We show the performance comparisons of all baselines and the proposed TLNet in Table 1. The DR-based methods consistently outperform the IPS-based methods and the propensity-independent methods, suggesting the strong ability of the DR method for debiasing and the necessity of guaranteeing the robustness of the debiasing model. TDR obtains the optimal performance among baselines due to the use of targeted learning. It highlights that targeted learning is useful to improve prediction performance. Nevertheless, TDR only imposes the target learning on the imputation model while we directly apply target learning to the ideal loss. Note that among all three datasets, the proposed TLNet consistently outperforms TDR and other baseline methods in AUC, NDCG@K, and Recall@K metrics. These findings underscore the robustness and superior performance of our proposed TLNet method in both ranking and retrieval tasks, demonstrating its strong potential for real-world applications in recommender systems.

Conclusions

This paper proposes the TLNet, a novel method that jointly train the propensity and prediction models and improve the efficiency and robustness of the prediction model in RS via the target learning. Specifically, we add a target function, which is the efficient influence curve of the ideal loss, to guide the optimization in the most efficient direction. Meanwhile, the proposed target function forces the bias of the loss to be sufficiently small, even under the misspecification of both the prediction and propensity models. TLNet possesses good properties such as robustness and efficiency inherited from the target learning. Experimental results on three real-world datasets demonstrate the superiority of TLNet.

References

Ai, Q.; Bi, K.; Luo, C.; Guo, J.; and Croft, W. B. 2018. Unbiased learning to rank with unbiased propensity estimation. In *SIGIR*.

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 *KDD*.

Ding, S.; Wu, P.; Feng, F.; He, X.; Wang, Y.; Liao, Y.; and Zhang, Y. 2022. Addressing Unmeasured Confounder for Recommendation with Sensitivity Analysis. In *KDD*.

Gao, C.; Li, S.; Lei, W.; Chen, J.; Li, B.; Jiang, P.; He, X.; Mao, J.; and Chua, T.-S. 2022. KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In *CIKM*.

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 *SIGIR*.

Kang, J. D.; and Schafer, J. L. 2007. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data.

Kweon, W.; and Yu, H. 2024. Doubly Calibrated Estimator for Recommendation on Data Missing Not At Random. In *WWW*.

Li, H.; Lyu, Y.; Zheng, C.; and Wu, P. 2023a. TDR-CL: Targeted Doubly Robust Collaborative Learning for Debiased Recommendations. In *ICLR*.

Li, H.; Xiao, Y.; Zheng, C.; Wu, P.; and Cui, P. 2023b. Propensity Matters: Measuring and Enhancing Balancing for Recommendation. In *ICML*.

Li, H.; Xiao, Y.; Zheng, C.; Wu, P.; Geng, Z.; Chen, X.; and Cui, P. 2024a. Debiased Collaborative Filtering with Kernel-based Causal Balancing. In *ICLR*.

Li, H.; Zheng, C.; Ding, S.; Feng, F.; He, X.; Geng, Z.; and Wu, P. 2024b. Be Aware of the Neighborhood Effect: Modeling Selection Bias under Interference for Recommendation. In *ICLR*.

Li, H.; Zheng, C.; Wang, S.; Wu, K.; Wang, E.; Wu, P.; Geng, Z.; Chen, X.; and Zhou, X.-H. 2024c. Relaxing the Accurate Imputation Assumption in Doubly Robust Learning for Debiased Collaborative Filtering. In *ICML*.

Li, H.; Zheng, C.; Wang, W.; Wang, H.; Feng, F.; and Zhou, X.-H. 2024d. Debiased recommendation with noisy feedback. In *KDD*.

Li, H.; Zheng, C.; and Wu, P. 2023. StableDR: Stabilized Doubly Robust Learning for Recommendation on Data Missing Not at Random. In *ICLR*.

Liu, D.; Cheng, P.; Zhu, H.; Dong, Z.; He, X.; Pan, W.; and Ming, Z. 2021. Mitigating Confounding Bias in Recommendation via Information Bottleneck. In *RecSys*.

Marlin, B.; Zemel, R. S.; Roweis, S.; and Slaney, M. 2007. Collaborative filtering and the missing at random assumption. In *RecSys*.

Marlin, B. M.; and Zemel, R. S. 2009. Collaborative prediction and ranking with non-random missing data. In *RecSys*.

Molenberghs, G.; Fitzmaurice, G.; Kenward, M. G.; Tsiatis, A.; and Verbeke, G. 2014. *Handbook of missing data methodology*. CRC Press.

Saito, Y. 2020. Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. In *SIGIR*.

Saito, Y. 2020. Doubly robust estimator for ranking metrics with post-click conversions. In *RecSys*.

Saito, Y.; Yaginuma, S.; Nishino, Y.; Sakata, H.; and Nakata, K. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *WSDM*.

Schnabel, T.; Swaminathan, A.; Singh, A.; Chandak, N.; and Joachims, T. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *ICML*, 1670–1679.

Seaman, S. R.; and Vansteelandt, S. 2018. Introduction to double robust methods for incomplete data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 33(2): 184.

Shi, C.; Blei, D. M.; and Veitch, V. 2019. Adapting Neural Networks for the Estimation of Treatment Effects. In *NeurIPS*.

Shi, Y.; Larson, M.; and Hanjalic, A. 2014. Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. *ACM Computing Surveys (CSUR)*, 47(1): 1–45.

Song, Z.; Chen, J.; Zhou, S.; Shi, Q.; Feng, Y.; Chen, C.; and Wang, C. 2023. CDR: Conservative doubly robust learning for debiased recommendation. In *CIKM*.

Steck, H. 2010. Training and testing of recommender systems on data missing not at random. In *KDD*.

Van der Laan, M. J.; Rose, S.; et al. 2011. *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer.

Vermeulen, K.; and Vansteelandt, S. 2015. Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, 110(511): 1024–1036.

Wang, F.; Zhong, W.; Xu, X.; Rafique, W.; Zhou, Z.; and Qi, L. 2020a. Privacy-aware cold-start recommendation based on collaborative filtering and enhanced trust. In *DSAA*.

Wang, F.; Zhu, H.; Srivastava, G.; Li, S.; Khosravi, M. R.; and Qi, L. 2021. Robust collaborative filtering recommendation with user-item-trust records. *TCSS*.

Wang, H.; Chang, T.-W.; Liu, T.; Huang, J.; Chen, Z.; Yu, C.; Li, R.; and Chu, W. 2022. Escm2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *SIGIR*.

Wang, L.; Ma, C.; Wu, X.; Qiu, Z.; Zheng, Y.; and Chen, X. 2024. Causally Debiased Time-aware Recommendation. In *WWW*.

Wang, W.; Zhang, Y.; Li, H.; Wu, P.; Feng, F.; and He, X. 2023. Causal recommendation: Progresses and future directions. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 3432–3435.

Wang, X.; Zhang, R.; Sun, Y.; and Qi, J. 2019. Doubly robust joint learning for recommendation on data missing not at random. In *ICML*.

Wang, Z.; Chen, X.; Wen, R.; Huang, S.-L.; Kuruoglu, E. E.; and Zheng, Y. 2020b. Information Theoretic Counterfactual Learning from Missing-Not-At-Random Feedback. In *NeurIPS*.

Xiao, Y.; Li, H.; Tang, Y.; and Zhang, W. 2024. Addressing Hidden Confounding with Heterogeneous Observational Datasets for Recommendation. In *NeurIPS*.

Yang, M.; Cai, G.; Liu, F.; Jin, J.; Dong, Z.; He, X.; Hao, J.; Shao, W.; Wang, J.; and Chen, X. 2023. Debiased recommendation with user feature balancing. *ACM Transactions on Information Systems*, 1–25.

Yang, M.; Dai, Q.; Dong, Z.; Chen, X.; He, X.; and Wang, J. 2021. Top-n recommendation with counterfactual user preference simulation. In *CIKM*.

Yuan, B.; Hsia, J.-Y.; Yang, M.-Y.; Zhu, H.; Chang, C.-Y.; Dong, Z.; and Lin, C.-J. 2019. Improving Ad Click Prediction by Considering Non-displayed Events. In *CIKM*.

Zhang, H.; Luo, F.; Wu, J.; He, X.; and Li, Y. 2023. LightFR: Lightweight federated recommendation with privacy-preserving matrix factorization. *ACM Transactions on Information Systems*, 1–28.

Zhang, H.; Wang, S.; Li, H.; Zheng, C.; Chen, X.; Liu, L.; Luo, S.; and Wu, P. 2024. Uncovering the Propensity Identification Problem in Debiased Recommendations. In *ICDE*.

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 *WWW*.

Zhang, Y.; Feng, F.; He, X.; Wei, T.; Song, C.; Ling, G.; and Zhang, Y. 2021. Causal intervention for leveraging popularity bias in recommendation. In *SIGIR*.