Reducing Item Discrepancy via Differentially Private Robust Embedding Alignment for Privacy-Preserving Cross Domain Recommendation

Weiming Liu [1] **Xiaolin Zheng*** [1] **Chaochao Chen** [1] **Jiahe Xu** [1] **Xinting Liao** [1] **Fan Wang** [1] **Yanchao Tan** [2] **Yew-Soon Ong** [3 4]

Abstract

Cross-Domain Recommendation (CDR) have become increasingly appealing by leveraging useful information to tackle the data sparsity problem across domains. Most of latest CDR models assume that domain-shareable user-item information (e.g., rating and review on overlapped users or items) are accessible across domains. However, these assumptions become impractical due to the strict data privacy protection policy. In this paper, we propose Reducing Item Discrepancy (`RidCDR`) model on solving Privacy-Preserving Cross-Domain Recommendation (PPCDR) problem. Specifically, we aim to enhance the model performance on both source and target domains without overlapped users and items while protecting the data privacy. We innovatively propose private-robust embedding alignment module in `RidCDR` for knowledge sharing across domains while avoiding negative transfer privately. Our empirical study on Amazon and Douban datasets demonstrates that `RidCDR` significantly outperforms the state-of-the-art models under the PPCDR without overlapped users and items.

1. Introduction

Recommender systems have become increasingly appealing in addressing the issue of information overload for users (Chen et al., 2023b; Zhu et al., 2021a; 2023; Zang et al., 2022; Li et al., 2024a; Liu et al., 2023b; Li et al., 2024c). Meanwhile, most of current Cross-Domain Recommendation (CDR) models assume the presence of prior-known overlapped users/items and the accessibility of intra-domain information (e.g., user-item ratings and reviews). However, these assumptions could be rather impractical with the increasing law on privacy protection (Pramod, 2023; Li et al., 2024c;b). In this scenario, user/item identities may undergo anonymous process and thus it could be hard to determine overlapped users or items (Zhang et al., 2021a). How to leverage cross-domain information while considering the aforementioned constraints remains a challenge for inclusive public service.

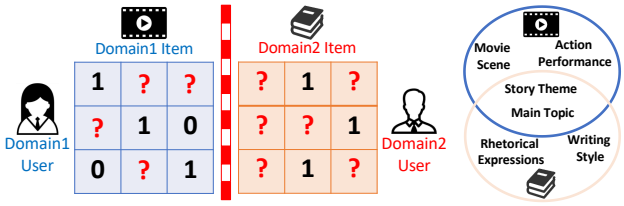

Figure 1. The illustration of privacy-preserving cross-domain recommendation problem without overlapped users and items.

In this paper, we focus on Privacy-Preserving Cross-Domain Recommendation (PPCDR) problem as shown in Fig.1. That is, we aim to resolve data sparsity problem and enhance the model performance on both source and target domains simultaneously. In each domain, user-item rating and review information is available, while other domains do not have access to these intra-domain contents. Moreover, we assume that both users and items are non-overlapped across domains which makes the problem even more challenging, especially under the settings of data privacy protection.

Previous CDR methods cannot better solve PPCDR well. On the one hand, conventional CDR and PPCDR models should utilize domain-shareable contents (e.g., overlapped users/items) as the bridges for knowledge sharing (Man et al., 2017; Cao et al., 2023; Zhu et al., 2022). These model performance could be severely degraded without the aid of overlapped users/items which ultimately limits their potentials in the real practice. On the other hand, items on different domains are inherently diverse and heterogeneous which leads to biases and discrepancies (Li et al., 2021). As the illustration in Fig.1., movie items inherit scene and action information while book items contain writing style. Meanwhile we should explore domain-invariant patterns (e.g.,

[1]Zhejiang University, China. [2]Fuzhou University, China. [3]Nanyang Technology University, Singapore. [4]Agency for Science, Technology and Research, Singapore. Correspondence to: Xiaolin Zheng <xlzheng@zju.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

theme and topic) among the movies and books to overcome the domain drift. Previous CDR models always utilize domain adversarial training strategy with discriminator (Ganin et al., 2016) for embedding adaptation in the latent space. However such approach could be unstable and difficult to train under certain circumstances (Shu et al., 2018). More recently, (Yang et al., 2023b) further adopt optimal transport method to align item embeddings for knowledge sharing. Nevertheless, traditional optimal transport could be easily affected by outliers and noise, thus leading to the negative transfer phenomenon (Mukherjee et al., 2021; Balaji et al., 2020; Le et al., 2021). In conclusion, how to reduce item domain shift and discrepancy without domain-shareable contents privately still need more investigation.

To address the aforementioned issues, in this paper, we propose Reducing Item Discrepancy model (`RidCDR`) for solving the PPCDR problem. `RidCDR` includes two modules, i.e., *rating prediction module* and *private-robust embedding alignment module*. Specifically, rating prediction module is set for collaborative filtering in the single domain. To further align relevant items with similar characteristics privately, we first propose private-robust embedding alignment module in `RidCDR`. For that purpose, we propose Differentially Private-Robust Adaptation (**DPRA**) method with *differentially private projection* and *robust reweighted sample adaptation* components. That is, we initially employ *sample weight adjustment mechanism* to filter out irrelevant items via unbalanced optimal transport privately. After that we utilize *sample reweighting optimal transport* to further measure and reduce domain discrepancy. By incorporating rating prediction module and private-robust embedding alignment module, we can leverage useful information across domains under the privacy-preserving constraints to enhance the model performance on PPCDR problem.

We summarize the main contributions of this paper as follows: (1) We propose a novel framework, i.e., `RidCDR`, for solving the PPCDR problem by combining intra-domain rating prediction and cross domain private-robust embedding alignment. (2) We introduce the private-robust embedding alignment module with the DPRA method, which incorporates differentially private projection and robust reweighted sample adaptation. This approach ensures reliable knowledge sharing across domains, even in the absence of overlapping users or items. (3) Extensive studies on Douban and Amazon datasets show that `RidCDR` significantly improves the state-of-the-art models in the PPCDR scenario.

2. Related Work

Cross Domain Recommendation. Cross-Domain Recommendation (CDR) mainly leverage source and target domains to tackle the data sparsity problem (Zhao et al., 2020a; Kang et al., 2019; Ju et al., 2024). Existing CDR models can be divided into two types, i.e., *domain-shareable* methods and *non-domain-shareable* methods. Specifically, domain-shareable methods rely on overlapped users or items as the bridges for knowledge sharing (Zhang et al., 2023; Zhao et al., 2023; Xie et al., 2022). Some domain-shareable methods also adopt cross connect network (Hu et al., 2018) or cross domain graph message propagation (Liu et al., 2020; Zhao et al., 2019) to improve the model performance. Non-domain-shareable methods tackle a more challenging scenario when there is a lack of overlapping users/items (Choi et al., 2022). Most non-domain-shareable models should utilize auxiliary information (e.g., user-item reviews) (Choi et al., 2022) for enhancing the results via Maximum Mean Discrepancy (Liu et al., 2022a) or adversarial training strategy (Hao et al., 2021). Recently, (Yang et al., 2023b) further adopt the balance optimal transport technique for latent embedding alignment to reduce domain shift. All these methods should assume that both source and target data are accessible during the training procedure. However, how to realize robust knowledge sharing for the Non-domain-shareable scenario privately still needs more exploration.

Privacy-Preserving Cross Domain Recommendation. With the increasing emphasis on data privacy protection, how to conduct CDR task privately has become a hot topic. (Chen et al., 2022) first adopts differential private publishing (Zhao et al., 2020b; Yang et al., 2023a) on the overlapped user for knowledge distillation from rich to sparse domain. (Liao et al., 2023) further adopts adversarial model (Xu et al., 2020) to generative fake user-item for modelling. Meanwhile (Chai et al., 2020; Meihan et al., 2022; Liu et al., 2023a; Yan et al., 2022; Chen et al., 2023a) start to utilize federated learning approach (McMahan et al., 2017) for protecting clients' privacy via distributed learning. However, these methods require prior identification of the overlapping users or items across domains. Meanwhile, achieving such requests can be challenging in privacy-preserving scenarios, which ultimately restricts their potential.

3. Methodology

3.1. Framework of `RidCDR`

In this section, we will introduce the model details of proposed `RidCDR`. There are two domains, i.e., source domain $\mathcal{S}$ and target domain $\mathcal{T}$ involves. We use $\mathcal{X}$ to indicate the domain index as $\mathcal{X} = \{(\mathrm{src}), (\mathrm{trg})\}$ for simplification. We assume each domain have $N_U^{\mathcal{X}}$ users, $N_V^{\mathcal{X}}$ items, and $\boldsymbol{\mathcal{R}}^{\mathcal{X}} \in \mathbb{R}^{N_U^{\mathcal{X}} \times N_V^{\mathcal{X}}}$ rating matrices to record user-item interactions. For the i-th user and j-th item in the $\mathcal{X}$-th domain, it consists of the tuples $(E_{U,i}^{\mathcal{X}}, E_{V,j}^{\mathcal{X}}, \mathcal{R}_{ij}^{\mathcal{X}}, H_{ij}^{\mathcal{X}})$. $E_{U,i}^{\mathcal{X}}$ and $E_{V,j}^{\mathcal{X}}$ denote the one-hot ID vector for the i-th user and j-th item respectively. $\mathcal{R}_{ij}^{\mathcal{X}}$ and $H_{ij}^{\mathcal{X}}$ denote the rating and review information respectively. Note that the users and items

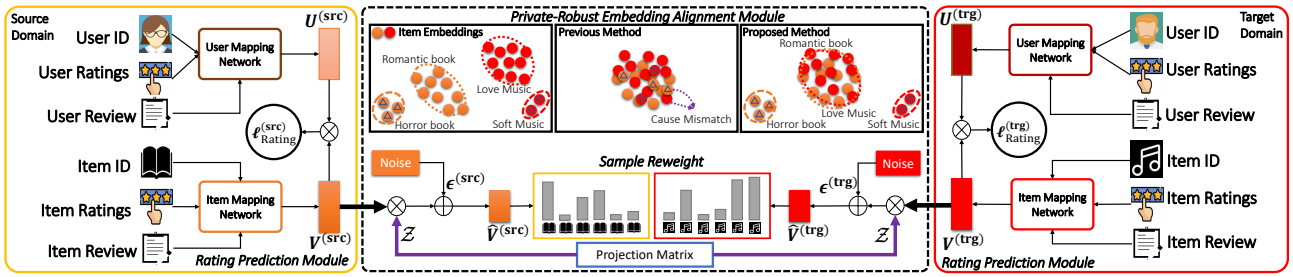

Figure 2. The framework of our proposed RidCDR.

are non-overlapped while user-item information (e.g, rating and review) cannot be directly accessible due to the privacy constraints. We wish to incorporate both source and target useful information for boosting the model performance.

Then, we introduce the overview of RidCDR, as is illustrated in Fig. 2. RidCDR mainly has two modules, i.e., *rating prediction module* and *private-robust embedding alignment module*. The rating prediction module is set to model user and item embeddings according to the interactions. The private-robust embedding alignment module aims to reduce the domain discrepancy among source and target domains while protecting the data privacy. By incorporating these two modules in RidCDR, we can leverage rich information from source domain to boost the model performance in the sparse target domain for knowledge sharing. We will introduce these two modules in detail later.

3.2. Rating Prediction Module

Firstly, we will introduce the rating prediction module for modelling intra-domain user-item embeddings. We adopt a trainable lookup table to exploit the user/item one-hot ID embedding as $\mathcal{E}^{\mathcal{X}}_{U,i} = \text{LookUp}(E^{\mathcal{X}}_{U,i})$ and $\mathcal{E}^{\mathcal{X}}_{V,j} = \text{LookUp}(E^{\mathcal{X}}_{V,j})$ respectively. For the i-th user and the j-th item, $\mathcal{R}^{\mathcal{X}}_{i*}$ and $\mathcal{R}^{\mathcal{X}}_{*j}$ denote the historical rating in domain $\mathcal{X}$. We utilize the fully connected layers $\mathcal{C}_{\mathcal{X}_U}$ and $\mathcal{C}_{\mathcal{X}_V}$ to obtain user and item behavior embeddings as $\mathcal{C}_{\mathcal{X}_U}(\mathcal{R}^{\mathcal{X}}_{i*}) = \mathcal{Q}^{\mathcal{X}}_{i*}$ and $\mathcal{C}_{\mathcal{X}_V}(\mathcal{R}^{\mathcal{X}}_{*j}) = \mathcal{Q}^{\mathcal{X}}_{*j}$. For the single review information $H^{\mathcal{X}}_{ij}$, we adopt BERT (Devlin et al., 2018) to obtain the review representations $\mathcal{H}^{\mathcal{X}}_{ij} = \text{BERT}(H^{\mathcal{X}}_{ij})$. Then we collect the user/item review representations by averaging the review representations given by the i-th users or the j-th item as $\mathcal{H}^{\mathcal{X}}_{i*}$ and $\mathcal{H}^{\mathcal{X}}_{*j}$ respectively. We also utilize the fully connected layers $\mathcal{K}_{\mathcal{X}_U}$ and $\mathcal{K}_{\mathcal{X}_V}$ to obtain user and item review embeddings as $\mathcal{K}_{\mathcal{X}_U}(\mathcal{H}^{\mathcal{X}}_{i*}) = \mathcal{O}^{\mathcal{X}}_{i*}$ and $\mathcal{K}_{\mathcal{X}_V}(\mathcal{H}^{\mathcal{X}}_{*j}) = \mathcal{O}^{\mathcal{X}}_{*j}$. We adopt the fully connected layer as user mapping network $F^{\mathcal{X}}_U(\cdot)$, $F^{\mathcal{X}}_V(\cdot)$ to obtain the user/item embedding for the $\mathcal{X}$-th domain respectively. That is, we concatenate one-hot ID embeddings, behavior embeddings, and review embeddings to obtain the user/item embedding. Specifically, the user embedding can be obtained via $F^{\mathcal{X}}_U(\mathcal{E}^{\mathcal{X}}_{U,i} \oplus \mathcal{Q}^{\mathcal{X}}_{i*} \oplus \mathcal{O}^{\mathcal{X}}_{i*}) = \boldsymbol{U}^{\mathcal{X}}_i \in \mathbb{R}^D$ where $\oplus$ denotes the concatenation operation and D denotes the embedding dimension. Likewise, we can obtain item embedding via $F^{\mathcal{X}}_V(\mathcal{E}^{\mathcal{X}}_{V,j} \oplus \mathcal{Q}^{\mathcal{X}}_{*j} \oplus \mathcal{O}^{\mathcal{X}}_{*j}) = \boldsymbol{V}^{\mathcal{X}}_j \in \mathbb{R}^D$. Finally, we adopt the multi-layer perceptron $G^{\mathcal{X}}(\cdot)$ to model the user-item ratings as $r^{\mathcal{X}}_{i,j} = G^{\mathcal{X}}(\boldsymbol{U}^{\mathcal{X}}_i \oplus \boldsymbol{V}^{\mathcal{X}}_j)$. We utilize the commonly-used binary cross-entropy loss as the rating prediction loss for collaborative filtering in domain $\mathcal{X}$ as:

$$\min \ell^{\mathcal{X}}_{\text{Rating}} = -\sum_{i,j}[\mathcal{R}^{\mathcal{X}}_{ij} \log r^{\mathcal{X}}_{i,j} + (1-\mathcal{R}^{\mathcal{X}}_{ij}) \log(1-r^{\mathcal{X}}_{i,j})].$$

Utilizing rating prediction module with optimizing $\ell^{\mathcal{X}}_{\text{Rating}}$ in both source and target domains, one can model user/item preferences using intra-domain data as shown in Fig.2.

3.3. Private-Robust Embedding Alignment Module

Although rating prediction module can model user-item interactions in the source or target domains, it could still result in limited performance due to the data sparsity problem. Therefore, it is essential to leverage useful dense source domain knowledge to improve the recommendation results in sparse target domain. Meanwhile we should notice that the original user/item embeddings cannot be directly share across domains due to the data privacy concerns. How to reduce domain discrepancy while protect data privacy become an urgent problem. To tackle the above issues, we will introduce the proposed private-robust embedding alignment module with newly proposed Differentially Private-Robust Adaptation (**DPRA**) method. Specifically, **DPRA** includes two main components, i.e., *differentially private projection* and *robust reweighted sample adaptation*. Differentially private projection component is set to enhance the data privacy during the domain adaptation process. Robust reweighted sample adaptation aims to further align similar item samples while avoiding negative transfer among dissimilar items.

3.3.1. Differentially Private Projection

Firstly, we introduce differentially private projection component. To strengthen data privacy, we adopt commonly-used differential privacy techniques (Abadi et al., 2016; Lê Tien et al., 2019a; Dwork, 2006; 2008) by adding the noise and projection into the original data.

Definition 1. (The definition of differential privacy (Abadi et al., 2016; Blocki et al., 2012; Lê Tien et al., 2019a)). *A randomized mechanism* $\mathcal{M} : \mathcal{X}^N \to \mathbb{R}^d$ *satisfied* (ϵ, δ)-

differential privacy if for any two datasets $X, X' \in \mathcal{X}^N$ *differing by a single element and for any set of possible output* $\mathcal{O} \subseteq \text{Range}(\mathcal{M})$*:*

$$\mathbb{P}(\mathcal{M}(X) \in \mathcal{O}) \leq e^{\epsilon}\mathbb{P}\left(\mathcal{M}\left(X'\right) \in \mathcal{O}\right) + \delta. \quad (1)$$

Specifically, we first generated a random projection matrix $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{D\times d}$ on both source and target domains. Meanwhile, we also sample noise matrix $\boldsymbol{\varepsilon}^{(\text{src})} \in \mathbb{R}^{N\times d}$ and $\boldsymbol{\varepsilon}^{(\text{trg})} \in \mathbb{R}^{N\times d}$ which generated from the normal distribution $\mathcal{N}(0, \sigma^2)$ respectively. σ denotes standard deviation and d denotes the projection dimension. Then we can transform origin item embeddings $\boldsymbol{V}^{(\text{src})}$ and $\boldsymbol{V}^{(\text{trg})}$ into privacy-preserving ones $\widehat{\boldsymbol{V}}^{(\text{src})}$ and $\widehat{\boldsymbol{V}}^{(\text{trg})}$ as:

$$\boldsymbol{V}^{(\text{src})}\boldsymbol{\mathcal{Z}} + \boldsymbol{\varepsilon}^{(\text{src})} = \widehat{\boldsymbol{V}}^{(\text{src})}, \boldsymbol{V}^{(\text{trg})}\boldsymbol{\mathcal{Z}} + \boldsymbol{\varepsilon}^{(\text{trg})} = \widehat{\boldsymbol{V}}^{(\text{trg})} \quad (2)$$

Note that we only publish projection matrix $\boldsymbol{\mathcal{Z}}$ and the noise matrix $\boldsymbol{\varepsilon}^{(\text{src})}$, $\boldsymbol{\varepsilon}^{(\text{trg})}$ will be only stored in the local domain.

Theorem 1. *Let* $\chi > 1$, $\delta \in [0, \frac{1}{2}]$ *and* $\boldsymbol{V} \in \mathbb{R}^{N\times D}$ *be the input data. Given a projection matrix* $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{D\times d}$ *generated from* $\mathcal{N}(0, \frac{1}{l})$ *and a noise matrix* $\boldsymbol{\varepsilon} \in \mathbb{R}^{N\times d}$ *generated from the normal distribution* $\mathcal{N}(0, \sigma^2)$*, the transformed data* $\widehat{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{\mathcal{Z}} + \boldsymbol{\varepsilon}$ *satisfied* (ϵ, δ)*-differential privacy. Specifically,* $\epsilon = \frac{\chi w(d,\delta/2)}{2\sigma^2} + \frac{\log(2/\delta)}{\chi - 1}$ *and* $w(d, \delta) = \frac{d}{D} + \frac{\Phi^{-1}(1-\delta)}{d}\sqrt{\frac{2d(D-1)}{d+2}}$ *where* $\Phi(\cdot)$ *denotes the cumulative distribution of distribution* $\mathcal{N}(0, \boldsymbol{I})$*.*

The proof of Theorem 1 can be found in (Rakotomamonjy & Liva, 2021). Based on Theorem 1, the transformed item embeddings $\widehat{\boldsymbol{V}}^{(\text{src})}$ and $\widehat{\boldsymbol{V}}^{(\text{trg})}$ satisfied (ϵ, δ)-differential privacy and thus they can protect data privacy (Kenthapadi et al., 2012; Lê Tien et al., 2019b; Blocki et al., 2012; Dwork et al., 2014; Wang et al., 2023). Thus we can conduct domain adaptation tasks among $\widehat{\boldsymbol{V}}^{(\text{src})}$ and $\widehat{\boldsymbol{V}}^{(\text{trg})}$.

3.3.2. Robust Reweighted Sample Adaptation

After we obtain the transformed item embeddings $\widehat{\boldsymbol{V}}^{(\text{src})}$ and $\widehat{\boldsymbol{V}}^{(\text{trg})}$, we should reduce the domain discrepancy for knowledge sharing. Traditional domain adaptation methods (e.g. balanced optimal transport (Villani et al., 2009)) always treat all samples equally. That is, all data samples should find a match during the calculation process. However, such methods can lead to negative transfer by aligning dissimilar items across domains. As illustrated in Fig.2, while romantic books and love music are well aligned, there is still a mismatch between horror books and soft music. Therefore, it is important while challenging to filter out these dissimilar items during domain adaptation for achieving more robustness results. To fulfill this task, we propose robust reweighted sample adaptation component with Unbalanced Optimal Transport (UOT) which includes two main steps, i.e., *Sample Weight Adjustment Mechanism* (SWAM) and *Sample Reweighting Optimal Transport* (SROT). SWAM is set to distinguish between items with similar or dissimilar properties via optimizing UOT across two domains. In the case of similar items, SWAM increases their sample weights while decreasing the sample weights for dissimilar items. SROT aims to further measure the discrepancy among the reweighted samples across domains. In this section, we will introduce the details of SWAM and SROT components.

Definition of UOT. To start with, we first illustrate the definition of UOT. Unlike balanced optimal transport, which imposes strict mass equality constraints across domains, UOT relaxes these constraints while improving feasibility in real-world applications (Benamou, 2003; Séjourné et al., 2022). Specifically, the formulation of UOT is given as:

$$\min_{\boldsymbol{\pi}\geq 0} J_{\text{UOT}} = \left[\langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau \text{KL}(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}) + \tau \text{KL}(\boldsymbol{\pi}^\top \mathbf{1}_N \| \boldsymbol{b})\right],$$

where $\boldsymbol{C} \in \mathbb{R}^{N\times N}$ denotes the cost matrix and it can be calculated via $C_{ij} = ||\widehat{\boldsymbol{V}}_i^{(\text{src})} - \widehat{\boldsymbol{V}}_j^{(\text{trg})}||_2^2$. $\boldsymbol{a}, \boldsymbol{b}$ denote the initial sample weights and we set them $a_i = b_j = \frac{1}{N}$ respectively. $\text{KL}(\boldsymbol{x}||\boldsymbol{y})$ denotes the KL Divergence between two d-dimensional data samples $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$ as $\text{KL}(\boldsymbol{x}||\boldsymbol{y}) = \sum_{i=1}^d \left[x_i \log \frac{x_i}{y_i} - x_i + y_i\right]$. Here τ denotes the balanced hyper parameter. Many previous papers involves entropy regularization term (Pham et al., 2020; Séjourné et al., 2019) or proximal term (Xie et al., 2020) into solving optimal $\boldsymbol{\pi}$ on UOT. However, it is worth noting that entropy regularization and proximal term can become numerically unstable during the iterations, leading to inaccurate matching solutions (Blondel et al., 2018). To tackle the optimization problem for J_{UOT}, we first propose SWAM to determine the importance on different data samples.

Optimization for SWAM. In this section, we will provide the technical details on solving J_{UOT} for data sample reweighting. Note that it could be difficult to directly optimize J_{UOT} without extra regularization terms (Pham et al., 2020). To fulfill the task, we should consider the dual form of J_{UOT}. That is, the Fenchel-Lagrange conjugate form of UOT with variables ζ, $\boldsymbol{u}$ and $\boldsymbol{v}$ as shown in Eq.(3):

$$\begin{aligned} &\min_{\boldsymbol{v},\boldsymbol{u},\zeta} L_{\text{UOT}} = \tau\left[e^{-\frac{\zeta}{\tau}}\langle \boldsymbol{a}, e^{-\frac{\boldsymbol{u}}{\tau}}\rangle + e^{\frac{\zeta}{\tau}}\langle \boldsymbol{b}, e^{-\frac{\boldsymbol{v}}{\tau}}\rangle\right], \\ &s.t.\ u_i + v_j \leq C_{ij}. \end{aligned} \quad (3)$$

Specifically, the source and target data samples can be reweighted through ζ, $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively:

$$\sum_{j=1}^N \pi_{ij} = a_i \exp\left(-\frac{u_i + \zeta}{\tau}\right), \quad \sum_{i=1}^N \pi_{ij} = b_j \exp\left(-\frac{v_j - \zeta}{\tau}\right).$$

The deduction details can be found in Appendix A. Thus we switch from optimizing transportation matrix $\boldsymbol{\pi}$ to data sample reweighting via optimizing variables ζ, $\boldsymbol{u}$ and $\boldsymbol{v}$.

Theorem 2. *Given the Fenchel-Lagrange conjugate form of UOT in Eq.(3), it guarantees the sum of sample weights are equal across domains. That is,* $\sum_{j=1}^N b_j \exp(-(v_j^* - \zeta^*)/\tau) = \sum_{i=1}^N a_i \exp(-(u_i^* + \zeta^*)/\tau)$ *where* ζ^*, $\boldsymbol{u}^*$ *and* $\boldsymbol{v}^*$ *denote the optimal solutions in Eq.(3) respectively.*

Proof. We first consider the Lagrange multipliers in Eq.(3):

$$\max_{\boldsymbol{\gamma}\geq 0}\min_{\boldsymbol{v},\boldsymbol{u},\zeta} L = L_{\mathrm{UOT}} + \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}(u_i + v_j - C_{ij}). \quad (4)$$

where $\boldsymbol{\gamma}$ is the Lagrange multipliers. Then we can take the differentiation w.r.t on $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively. We will obtain the results $\sum_{j=1}^{N}\gamma_{ij} = a_i \exp(-(u_i^* + \zeta^*)/a_i)$ and $\sum_{i=1}^{N}\gamma_{ij} = b_j \exp(-(v_j^* - \zeta^*)/b_j)$ accordingly. Therefore we can validate the correctness of Theorem 2.

Meanwhile we should also optimize multiple variables ζ, $\boldsymbol{u}$ and $\boldsymbol{v}$ for sample reweighting. To further simplify the optimization process, we adopt c-transform strategy (An et al., 2022) to figure out the upper bound of $\boldsymbol{v}$ as $v_j = \inf_{k\in[N]}(C_{kj} - u_k)$ by minimizing Eq.(3) while maintaining the constraint of $u_i + v_j \leq C_{ij}$. At that time, we only need to exploit the value of $\boldsymbol{u}$ and ζ as below:

$$\min_{\boldsymbol{u},\zeta} L_{\mathrm{U}} = \tau\left[e^{-\frac{\zeta}{\tau}}\sum_{i=1}^{N} a_i e^{-\frac{u_i}{\tau}} + e^{\frac{\zeta}{\tau}}\sum_{j=1}^{N} b_j e^{-\frac{\inf_{k\in[N]}(C_{kj}-u_k)}{\tau}}\right].$$

To optimize L_{U}, we first set $\zeta^{(0)} = 0$ for initialization. Then we fix $\zeta^{(t-1)}$ to optimize $\boldsymbol{u}^{(t)}$ via FISTA method (Kim & Fessler, 2018) at the t-th iteration. We calculate the sub-gradient (Boyd & Vandenberghe, 2004) w.r.t $\boldsymbol{u}$ on L_{U} as:

$$\mathcal{G}(u_i^{(t)}) = -e^{\frac{\zeta^{(t-1)}}{\tau}}\sum_{j=1}^{N} b_j e^{-\frac{\inf_{k\in[N]}(C_{kj}-u_k^{(t)})}{\tau}}\cdot\mathbb{I}(k^*=i) - a_i e^{-\frac{u_i^{(t)}+\zeta^{(t-1)}}{\tau}},$$

where $\mathbb{I}(\cdot)$ is the indicator function, which takes the value 1 when the condition inside is true, and 0 otherwise. After that, we could adopt Armijo line-search strategy (de Oliveira & Takahashi, 2021) to determine the step-size η on $\nabla_{u_i} L_{\mathrm{U}}$ for the following procedure. Then we could obtain the optimal solution on $\boldsymbol{u}^{(t)}$ and $\boldsymbol{v}^{(t)} = \inf_{k\in[N]}(C_{kj} - u_k^{(t)})$, we further optimize $\zeta^{(t)}$ by considering $\nabla_{\zeta^{(t)}} L_{\mathrm{U}} = 0$. Specifically, we can obtain the exact value of $\zeta^{(t)}$ via:

$$\zeta^{(t)} = \frac{\tau}{2}\log\left\langle \boldsymbol{a}, e^{-\frac{\boldsymbol{u}^{(t)}}{\tau}}\right\rangle - \frac{\tau}{2}\log\left\langle \boldsymbol{b}, e^{-\frac{\boldsymbol{v}^{(t)}}{\tau}}\right\rangle. \quad (5)$$

Apparently, we can easily verify that $\sum_{j=1}^{N} b_j \exp(-(v_j^{(t)} - \zeta^{(t)})/\tau) = \sum_{i=1}^{N} a_i \exp(-(u_i^{(t)} + \zeta^{(t)})/\tau)$ are satisfied during each iterations. It also highlights the significance of ζ in the optimization process. We can achieve the optimal solution ζ^*, $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ after several iterations. Moreover, we can observe that the KL-Divergence of $\mathrm{KL}(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a})$ and $\mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_N\|\boldsymbol{b})$ turns out to be constants once it converges to optimum:

$$\begin{cases}\mathrm{KL}(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}) = \sum_{i=1}^{N}\left[\left(1 - \frac{u_i^*+\zeta^*}{\tau}\right)a_i e^{-\frac{u_i^*+\zeta^*}{\tau}} + a_i\right]\\ \mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_N\|\boldsymbol{b}) = \sum_{j=1}^{N}\left[\left(1 - \frac{v_j^*-\zeta^*}{\tau}\right)b_j e^{-\frac{v_j^*-\zeta^*}{\tau}} + b_j\right].\end{cases}$$

Then we can find that the original unbalanced optimal transport problem has been transformed into classic optimal transport problem. Specifically, we can obtain the Sample Reweighting Optimal Transport (SROT) as follows:

$$\min_{\boldsymbol{\pi}\geq 0} J_{\mathrm{SROT}} = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau\cdot\mathrm{Constant}$$

$$s.t.\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}\odot e^{-\frac{\boldsymbol{u}^*+\zeta^*}{\tau}} = \widehat{\boldsymbol{a}},\ \boldsymbol{\pi}^\top\mathbf{1}_N = \boldsymbol{b}\odot e^{-\frac{\boldsymbol{v}^*-\zeta^*}{\tau}} = \widehat{\boldsymbol{b}},$$

where $\widehat{\boldsymbol{a}}$ and $\widehat{\boldsymbol{b}}$ denote the update weights for source and target samples respectively. That is, we reweight the samples with coefficients $e^{-\frac{\boldsymbol{u}^*+\zeta^*}{\tau}}$ or $e^{-\frac{\boldsymbol{v}^*-\zeta^*}{\tau}}$ accordingly. Samples that are more similar between the source domain and the target domain receive higher weights. Meanwhile, the weights of dissimilar samples will be reduced. Therefore, dissimilar items will not get matched or transported to avoid negative transfer, leading to more robust solutions. However, we still cannot reach the exact value on domain discrepancy. Hence we should further introduce the optimization on SROT for domain adaptation and knowledge sharing.

Algorithm 1 The optimization procedure on SWAM

Input: $\boldsymbol{C}$: The cost distance matrix; τ: Hyper parameters; $\boldsymbol{a}, \boldsymbol{b}$: Given the source and target marginal probabilities.
Procedure:

```
Initialize t = 0, u^(0) = (0, 0, ··· , 0).
repeat
    Update u^(t+1) = FISTA(u^(t), L_U, G(u^(t))).
    Update v^(t+1) = inf_{k=1}^N (C_kj − u_k^(t+1)).
    Update ζ^(t+1) using u^(t+1) and v^(t+1) via Eq.(5).
    Update t = t + 1.
until Converge
Function: FISTA (u^(t), L_U, G(u^(t))):
Initialize θ_0 = 1 and t' = 0.
repeat
    Calculate the sub-gradient G(u^(t)).
    η = LineSearch(u^(t), L_U, G(u^(t))).
    Update û^(t) = u^(t) − ηG(u^(t)).
    Update θ_{t'+1} = 1/2 (1 + sqrt(1 + 4θ_{t'}^2))
    Update u^(t) ← u^(t) + (θ_{t'} − 1)/θ_{t'+1} (û^(t) − u^(t)).
    Update t' = t' + 1.
until Converge.
Return: The optimal ζ*, u* and v*.
```

Optimization for SROT. To obtain the optimal result of J_{SROT} in Eq.(6), we first consider the dual Fenchel-Lagrange conjugate form of SROT as follows:

$$\max_{\boldsymbol{f},\boldsymbol{g}} \ell_{\mathrm{SROT}} = \langle \boldsymbol{f}, \widehat{\boldsymbol{a}}\rangle + \langle \boldsymbol{g}, \widehat{\boldsymbol{b}}\rangle\ s.t.\ f_i + g_j \leq C_{ij}. \quad (6)$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ denote the Lagrange multipliers. Specifically, we can set $g_j = \inf_{k=1}^{N}(C_{kj} - f_k)$ according to the c-transform theorem (Liu et al., 2023c). By taking c-transform back into Eq.(6), we can achieve the *Reweighting Discrete Kantorovich Functional (RDKF)* as follows:

$$\max_{\boldsymbol{f}\in\boldsymbol{\Delta}} \ell_{\mathrm{RDKF}} = \sum_{i=1}^{N} f_i\widehat{a}_i - \sum_{j=1}^{N}\widehat{b}_j\sup_{k\in[N]}(f_k - C_{kj}). \quad (7)$$

Note that we only need to optimize $\boldsymbol{f}$ during the whole optimization process. Meanwhile we also name $\boldsymbol{f}$ as *reweighting Kantorovich potential*. Since Eq.(7) could have multiple solutions (An et al., 2022), we further add zero-mean constraints on $\boldsymbol{f}$ as $\Delta = \{\sum_{i=1}^{N} f_i = 0\}$. Likewise, we further adopt FISTA algorithm to optimize Eq.(17) for finding the optimal solution on $\boldsymbol{f}^*$. The optimization details can be found in Appendix.B. After we obtain $\boldsymbol{f}^*$, the value of ℓ_{RDKF} could better reflect the discrepancy across domains.

3.4. Putting Together

The total loss of RidCDR could be obtained by combining the losses of rating prediction loss $\ell^{\mathcal{X}}_{\text{Rating}}$ and the robust reweighted sample adaptation loss ℓ_{RDKF} as follows:

$$\min \ell^{\mathcal{X}}_{\text{RidCDR}} = \ell^{\mathcal{X}}_{\text{Rating}} + \lambda \ell_{\text{RDKF}}, \tag{8}$$

where $\mathcal{X} = \{(\text{src}), (\text{trg})\}$ denotes the source and target domains. λ denotes the hyper parameter to balance these two kinds of losses. The optimization process are provided in Alg.2. In summary, we first adopt rating prediction module to model user and item embeddings. Then we adopt differentially private mechanism to protect the data privacy in private-robust embedding alignment module. Note that there is no need to adopt differential privacy in the rating prediction module, since the computation is within each domain and other domains will not directly get access to these data. Once we need to transfer knowledge across domains, therefore, we make projection and add noise to hide original item embeddings to make the projected item embeddings $\widehat{\boldsymbol{V}}^{(\text{src})}$ and $\widehat{\boldsymbol{V}}^{(\text{trg})}$ to obey (ϵ, δ)-differential privacy. By doing this procedure, we can prevent external attackers from intercepting the original item information. Meanwhile, other domains cannot restore the original item embeddings and causes the data leakage. Moreover, differential privacy inherits the post-processing property (Dwork et al., 2006; Qu et al., 2024; Wu et al., 2022; Qi et al., 2020), thus the subsequent computing of privacy-preserved data is also privacy-preserving. Finally we reweight source and target data samples and filter out irrelevant ones accordingly for figuring out the exact discrepancy across domains. Thus the useful knowledge can be transferred while the intra-domain information can be protected.

Algorithm 2 The training procedure of RidCDR

Procedure:

for epoch $= 1$ to K **do**
  Sample N number of tuples $(E^{\mathcal{X}}_{U,i}, E^{\mathcal{X}}_{V,j}, \mathcal{R}^{\mathcal{X}}_{ij}, H^{\mathcal{X}}_{ij})$.
  Obtain $\boldsymbol{U}^{\mathcal{X}}, \boldsymbol{V}^{\mathcal{X}}$ in domain $\mathcal{X}$.
  Minimize $L^{\mathcal{X}}_{R}$ in domain $\mathcal{X}$ for rating prediction.
  # Conducting DRPA method (Line 6-9)
  Sample matrix $\mathcal{Z}$ and noise $\boldsymbol{\epsilon}^{\mathcal{X}}$ in domain $\mathcal{X}$.
  Transform item embeddings $\boldsymbol{V}^{\mathcal{X}}\boldsymbol{\mathcal{Z}} + \boldsymbol{\epsilon}^{\mathcal{X}} = \widehat{\boldsymbol{V}}^{\mathcal{X}}$.
  Obtain $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^* in SWAM for data reweighting.
  Optimize $\boldsymbol{f}^*$ in ℓ_{RDKF} for SROT.
  Minimize the total loss function $\ell^{\mathcal{X}}_{\text{RidCDR}}$.
end for
Return: The well-trained recommendation model.

4. Experiments

In this section, we conduct experiments on several real-world datasets to evaluate our proposed models.

4.1. Experimental setup

Datasets. We conduct extensive experiments on two popularly used real-world datasets, i.e., *Amazon* and *Douban*. The **Amazon** dataset (Ni et al., 2019) has five datasets, i.e., Movie, Book, CD, Video, and Game. The **Douban** dataset (Zhu et al., 2019; 2021b) has three domains, i.e., Book, Music, and Movie. The detailed statistics of these datasets are given in Appendix.C. For each dataset, we binarize the ratings higher or equal to 4 as positive. We also filter the users and items with less than 5 interactions, following existing research (Zhu et al., 2019; Liu et al., 2022b).

Baseline. We compare our proposed **RidCDR** with the following state-of-the-art models. (1) **NeuMF** (He et al., 2017) is the most famous single-domain model which adopts neural network for collaborative filtering. (2) **DeepCoNN** (Zheng et al., 2017) fuses ID information and user textual features for recommendation. (3) **NARRE** (Chen et al., 2018) utilizes CNN with attention module to extract review-level information for user modelling. (4) **DDTCDR** (Li & Tuzhilin, 2020) utilizes orthogonal transformation between source and target user embeddings. (5) **Rec-DAN** (Wang et al., 2019a) aligns user-item textual features via the adversarial training strategy for transferring useful knowledge. (6) **CGN** (Zhang et al., 2021b) adopts a cycle generation networks with dual-direction mapping for personalized CDR tasks (7) **TDAR** (Yu et al., 2020) utilizes adversarial training strategy on review text memory network for cross-domain recommendation. (8) **CDRIB** (Cao et al., 2022) adopts variational information bottleneck to exploit user preferences. (9) **DisAlign** (Liu et al., 2021) adopts Stein path with probability estimation for reducing domain discrepancy. (10) **CFAA** (Liu et al., 2022a) adopts the horizontal and vertical attribute alignment for domain adaptation on recommendation. (11) **GWCDR** (Li et al., 2022) carries out Gromov-Wasserstein distance on graph matching for knowledge sharing. (12) **DURation** (Lu et al., 2022) adopts distribution variance and correlation alignment to obtain unified representations. (13) **SER** (Choi et al., 2022) considers review-based domain disentanglement model without duplicate users for cross-domain recommendation. (14) **SRTrans** (Li et al., 2023) clusters ratings and reviews for cross-domain knowledge sharing. (15) **MOTKD** (Yang et al., 2023b) utilizes classic optimal transport with proxy

	(Amazon) CD → Movie				(Amazon) Movie → CD				(Amazon) Video → Game			
	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20
NeuMF	0.2773	0.1069	0.4540	0.1535	0.3317	0.1426	0.5292	0.1864	0.2781	0.1093	0.4628	0.1506
DeepCoNN	0.2910	0.1142	0.4729	0.1597	0.3466	0.1504	0.5438	0.2021	0.2853	0.1267	0.4805	0.1635
NARRE	0.2982	0.1257	0.4813	0.1754	0.3528	0.1589	0.5576	0.2105	0.2960	0.1343	0.4872	0.1711
DDTCDR	0.3154	0.1336	0.4957	0.1840	0.3703	0.1792	0.5711	0.2289	0.3225	0.1572	0.5004	0.1868
Rec-DAN	0.3249	0.1383	0.5024	0.1928	0.3774	0.1851	0.5786	0.2372	0.3331	0.1659	0.5165	0.2007
CGN	0.3331	0.1450	0.5115	0.1963	0.3892	0.1946	0.5827	0.2485	0.3612	0.1634	0.5279	0.2153
TDAR	0.3428	0.1544	0.5093	0.1971	0.3923	0.1929	0.5902	0.2466	0.3650	0.1744	0.5335	0.2182
CDRIB	0.3576	0.1590	0.5271	0.2113	0.3969	0.2007	0.5942	0.2535	0.3824	0.1820	0.5478	0.2261
DisAlign	0.3465	0.1637	0.5323	0.2152	0.4015	0.1970	0.6034	0.2494	0.3886	0.1859	0.5551	0.2302
CFAA	0.3617	0.1694	0.5362	0.2239	0.4076	0.2042	0.6108	0.2655	0.3937	0.1931	0.5610	0.2353
GWCDR	0.3642	0.1755	0.5538	0.2196	0.4024	0.2083	0.6061	0.2719	0.4045	0.2027	0.5664	0.2473
DURation	0.3683	0.1819	0.5621	0.2344	0.4178	0.2165	0.6127	0.2752	0.4023	0.2078	0.5715	0.2509
SER	0.3764	0.1853	0.5670	0.2382	0.4259	0.2231	0.6205	0.2746	0.4095	0.2164	0.5783	0.2571
SRTrans	0.3730	0.1861	0.5714	0.2425	0.4332	0.2296	0.6273	0.2807	0.4138	0.2219	0.5857	0.2620
MOTKD	0.3856	0.1902	0.5789	0.2467	0.4431	0.2348	0.6350	0.2845	0.4204	0.2275	0.5936	0.2668
RidCDR-B	0.3067	0.1304	0.4892	0.1815	0.3641	0.1655	0.5637	0.2163	0.3119	0.1428	0.4942	0.1776
RidCDR-D	0.3904	0.2035	0.5866	0.2451	0.4287	0.2264	0.6229	0.2751	0.4174	0.2237	0.5852	0.2680
RidCDR-U	0.3991	0.2076	0.5915	0.2503	0.4460	0.2375	0.6433	0.2873	0.4243	0.2306	0.5987	0.2727
RidCDR	**0.4088**	**0.2228**	**0.5977**	**0.2590**	**0.4672**	**0.2477**	**0.6664**	**0.2980**	**0.4283**	**0.2340**	**0.6114**	**0.2768**
	(Amazon) CD → Book				(Amazon) Book → CD				(Amazon) Game → Video			
	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20
NeuMF	0.1729	0.0608	0.3571	0.1062	0.3081	0.1336	0.5128	0.1875	0.2414	0.0793	0.4082	0.1430
DeepCoNN	0.1938	0.0744	0.3726	0.1251	0.3272	0.1423	0.5215	0.2004	0.2590	0.0986	0.4207	0.1559
NARRE	0.2040	0.0823	0.3819	0.1316	0.3403	0.1495	0.5337	0.2086	0.2654	0.1132	0.4261	0.1628
DDTCDR	0.2212	0.0902	0.3983	0.1470	0.3546	0.1638	0.5459	0.2241	0.2803	0.1218	0.4471	0.1754
Rec-DAN	0.2377	0.0969	0.4108	0.1525	0.3632	0.1740	0.5584	0.2335	0.2865	0.1270	0.4658	0.1704
CGN	0.2525	0.1040	0.4247	0.1663	0.3715	0.1866	0.5731	0.2472	0.3039	0.1384	0.4816	0.1823
TDAR	0.2568	0.1073	0.4319	0.1724	0.3672	0.1785	0.5763	0.2414	0.2986	0.1351	0.4792	0.1776
CDRIB	0.2709	0.1225	0.4434	0.1756	0.3963	0.1911	0.5875	0.2547	0.3114	0.1482	0.4938	0.1921
DisAlign	0.2734	0.1262	0.4459	0.1788	0.3971	0.1965	0.5947	0.2523	0.3152	0.1513	0.5007	0.1941
CFAA	0.2771	0.1324	0.4510	0.1832	0.4065	0.2016	0.5998	0.2561	0.3138	0.1547	0.5091	0.2003
GWCDR	0.2956	0.1431	0.4642	0.1885	0.4097	0.2064	0.6053	0.2588	0.3210	0.1626	0.5173	0.2059
DURation	0.3043	0.1475	0.4691	0.1906	0.4150	0.2132	0.6114	0.2615	0.3287	0.1695	0.5229	0.2080
SER	0.3017	0.1526	0.4758	0.1963	0.4144	0.2157	0.6039	0.2662	0.3351	0.1729	0.5244	0.2105
SRTrans	0.3096	0.1573	0.4825	0.2044	0.4189	0.2203	0.6112	0.2638	0.3389	0.1774	0.5310	0.2166
MOTKD	0.3120	0.1699	0.4784	0.2021	0.4253	0.2223	0.6148	0.2720	0.3455	0.1801	0.5367	0.2212
RidCDR-B	0.2085	0.0876	0.3922	0.1369	0.3471	0.1534	0.5410	0.2153	0.2742	0.1197	0.4405	0.1674
RidCDR-D	0.3154	0.1618	0.4835	0.2092	0.4237	0.2246	0.6186	0.2737	0.3446	0.1766	0.5359	0.2190
RidCDR-U	0.3283	0.1757	0.4923	0.2177	0.4282	0.2289	0.6202	0.2770	0.3565	0.1825	0.5436	0.2275
RidCDR	**0.3359**	**0.1787**	**0.5007**	**0.2199**	**0.4384**	**0.2343**	**0.6299**	**0.2826**	**0.3627**	**0.1903**	**0.5645**	**0.2413**
	(Douban) Movie				(Douban) Book				(Douban) Music			
	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20	HR@10	NDCG@10	HR@20	NDCG@20
NeuMF	0.1936	0.0875	0.2693	0.1288	0.2451	0.1067	0.3001	0.1544	0.3115	0.1396	0.3478	0.1650
DeepCoNN	0.2281	0.0997	0.2805	0.1364	0.2616	0.1229	0.3175	0.1760	0.3202	0.1483	0.3690	0.1752
NARRE	0.2407	0.1113	0.2974	0.1459	0.2762	0.1281	0.3393	0.1825	0.3277	0.1542	0.3765	0.1834
DDTCDR	0.2569	0.1254	0.3082	0.1592	0.2838	0.1373	0.3541	0.1885	0.3556	0.1709	0.3984	0.1910
Rec-DAN	0.2775	0.1324	0.3191	0.1676	0.2901	0.1439	0.3659	0.1962	0.3623	0.1797	0.4082	0.1965
CGN	0.2858	0.1439	0.3240	0.1734	0.2957	0.1462	0.3686	0.2013	0.3671	0.1825	0.4163	0.2031
TDAR	0.2834	0.1395	0.3378	0.1754	0.2980	0.1483	0.3614	0.1979	0.3688	0.1862	0.4141	0.2086
CDRIB	0.2922	0.1557	0.3446	0.1710	0.3165	0.1524	0.3772	0.2026	0.3739	0.1853	0.4180	0.2169
DisAlign	0.2956	0.1531	0.3487	0.1762	0.3193	0.1557	0.3763	0.2070	0.3765	0.1944	0.4239	0.2212
CFAA	0.2998	0.1586	0.3522	0.1831	0.3254	0.1618	0.3879	0.2146	0.3750	0.1893	0.4305	0.2268
GWCDR	0.3067	0.1610	0.3573	0.1894	0.3226	0.1645	0.3837	0.2169	0.3822	0.2001	0.4292	0.2317
DURation	0.3139	0.1673	0.3648	0.1955	0.3316	0.1697	0.3924	0.2201	0.3887	0.2075	0.4424	0.2335
SER	0.3185	0.1644	0.3590	0.1868	0.3279	0.1732	0.3913	0.2087	0.3866	0.2021	0.4342	0.2380
SRTrans	0.3261	0.1708	0.3712	0.2003	0.3397	0.1776	0.3980	0.2232	0.3915	0.2064	0.4373	0.2329
MOTKD	0.3353	0.1769	0.3781	0.2077	0.3462	0.1834	0.4026	0.2290	0.3945	0.2128	0.4411	0.2376
RidCDR-B	0.2475	0.1164	0.3021	0.1502	0.2793	0.1348	0.3530	0.1867	0.3469	0.1655	0.3926	0.1889
RidCDR-D	0.3314	0.1735	0.3766	0.2050	0.3478	0.1809	0.4081	0.2323	0.3943	0.2072	0.4459	0.2407
RidCDR-U	0.3390	0.1782	0.3833	0.2126	0.3521	0.1877	0.4135	0.2389	0.4004	0.2215	0.4509	0.2441
RidCDR	**0.3572**	**0.1836**	**0.3905**	**0.2184**	**0.3639**	**0.1953**	**0.4218**	**0.2462**	**0.4131**	**0.2290**	**0.4587**	**0.2525**

Table 1. Experimental results on Douban and Amazon datasets.

	(Amazon) Video → Game			
	HR@10	NDCG@10	HR@20	NDCG@20
RidCDR	0.4283	0.2340	0.6114	0.2768
RidCDR++	**0.4369**	**0.2385**	**0.6176**	**0.2820**
	(Amazon) Game → Video			
	HR@10	NDCG@10	HR@20	NDCG@20
RidCDR	0.3627	0.1903	0.5645	0.2416
RidCDR++	**0.3678**	**0.1966**	**0.5723**	**0.2474**

Table 2. Method extension results on Amazon datasets.

distributions for feature alignment across domains which is the state-of-the-art cross-domain recommendation model.

Implemented details. We provide the implemented details of our proposed RidCDR. We set batch size $N = 256$ and embedding dimension as $D = 128$ across different domains. We set $d = \frac{D}{2} = 64$ and $\delta = \frac{1}{1.2N}$ according to the previous experimental protocol (Papernot et al., 2016). Then we set the privacy budget $\epsilon = 4$ in differentially private projection module. Note that σ can be obtained using the RDP based moment accountant following (Rakotomamonjy & Liva, 2021). We set $\tau = 0.5$ for SWAM in robust reweighted sample adaptation. Finally we set $\lambda = 0.1$ for ℓ_{RDKF} on the total loss of RidCDR. We choose Adam (Kingma & Ba, 2014; Zhu et al., 2020) as optimizer, and adopt Hit Rate (HR), and NDCG (Wang et al., 2019b) as the evaluation metrics. For all the experiments, we perform five random experiments and report the average results. We report the results measured by the commonly used metrics as $\{\mathrm{Top@10}, \mathrm{Top@20}\}$ in both Douban and Amazon datasets.

4.2. Recommendation Performance

The comparison results on Douban and Amazon datasets are shown in Table 1. From it, we can observe that: (1) Only utilizing single domain information (e.g., NeuMF and DeepCoNN) cannot achieve better results since they fail to resolve the data sparsity problem. (2) Involving cross domain information could improve the model performance by

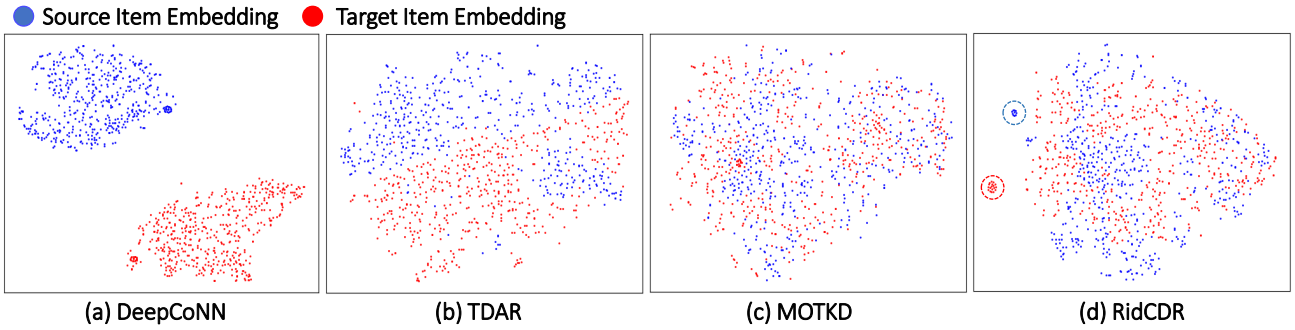

Figure 3. The T-SNE plot on item embeddings (Amazon Movie and Amazon CD) for DeepCoNN, TDAR, MOTKD, and RidCDR.

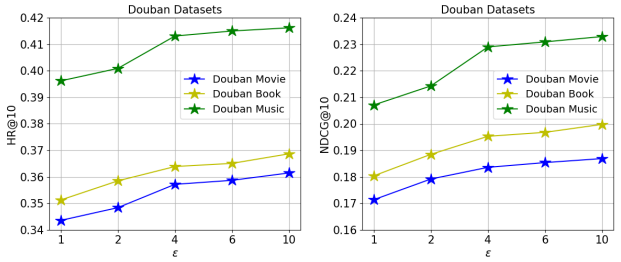

Figure 4. The experimental results on varying ϵ.

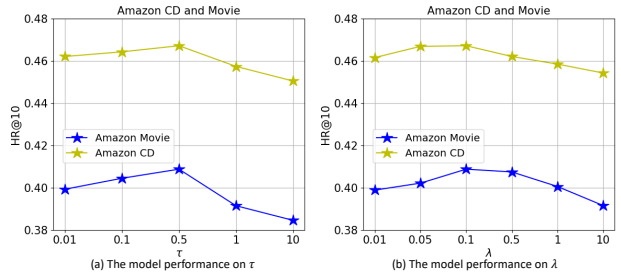

Figure 5. The experimental results on varying τ and λ.

knowledge sharing. However, conventional methods (e.g., DDTCDR and GWCDR) cannot fully explore useful textual representations and limits their potentials. (3) Adopting both rating and review for collaborative filtering can even enhance the output results. Nevertheless, current models (e.g., TDAR and SER) usually adopt adversarial training strategy which could be unstable (Shu et al., 2018) and hard to train in real practice. (4) Our proposed RidCDR achieves the best performance which indicating the efficacy of DPRA method for alleviating domain discrepancy in private-robust embedding alignment module. Moreover, RidCDR with differentially private projection can further protect data privacy during the training procedure. Furthermore, we also adopt T-SNE (Van der Maaten & Hinton, 2008) to visualize the latent item embeddings among Amazon Movie and CD as shown in Fig.3(a)-(d). We can find that the domain discrepancy are exist in Fig.3(a) while DeepCoNN cannot better resolve. Thus single-domain recommendation models cannot realize knowledge sharing for tackling the data sparsity problem. Although TDAR attempts to narrow the gap between the source and target domains using feature discriminators, it is still evident that the embeddings have not been effectively aligned together in Fig.3(b). MOTKD further adopts standard optimal transport for embedding adaptation with better results than TDAR in Fig.3(c). Moreover, our proposed RidCDR can even filter out irrelevant samples in Fig.3(d), thereby avoiding negative transfer in MOTKD for enhancing the model's representation capability.

4.3. Analysis

Ablation study. To study how does each module of **RidCDR** contribute on the final performance, we compare **RidCDR** with its several variants, including **RidCDR-B**, **RidCDR-D**, and **RidCDR-U** on Douban and Amazon datasets. Specifically, **RidCDR-B** only utilizes rating prediction module for the recommendation. **RidCDR-D** and **RidCDR-U** adopt DP-Sliced OT (Rakotomamonjy & Liva, 2021) and entropic UOT (Pham et al., 2020) for aligning item embeddings across domains. From that we can observe: (1) **RidCDR-B** cannot provide satisfied results since it cannot share useful knowledge among source and target domains to alleviate the data sparsity problem. (2) **RidCDR-D** and **RidCDR-U** achieve much better results than **RidCDR-B**, indicating that it is essential to reduce domain bias and discrepancy. However, **RidCDR-D** cannot filter out irrelevant items and thus leads to the negative transfer. **RidCDR-U** with entropy regularization term could lead to inaccurate mapping solution (Blondel et al., 2018) which hurdle the model performance. (3) Comparing RidCDR-D, RidCDR-U, and RidCDR, we conclude that utilizing DPRA for private-robust embedding alignment can boost the model potentials.

Method Extension. We also investigate the method extension on our proposed RidCDR. That is, we conduct the differentially private-robust adaptation method on both users and items embeddings as RidCDR++. We conduct the experiments on Amazon Game and Video and report the results on Table.2. From this observation, it can be noted that RidCDR++ not only slightly improves the model performance compared to RidCDR, but also suggests that the proposed method can be effectively and privately employed for knowledge sharing on both users and items.

Parameter sensitivity. We further study the effects of

hyper-parameters on model performance. We first vary ϵ in $\epsilon \in \{1, 2, 4, 6, 10\}$ for the analysis on differential privacy on Douban datasets. We report the results of HR@10 and NDCG@10 as shown in Fig.4. We can observe that our proposed RidCDR can be suitable for different kinds of privacy budgets to obtain satisfying performance. That is, larger ϵ value could yield superior results, while a smaller ϵ value can provide better protection for data privacy. To balance the data privacy protection capacity and the model performance, we set it as $\epsilon = 4$ empirically. We also vary τ in $\tau \in \{0.05, 0.1, 0.5, 1, 10\}$ on Amazon CD and Amazon Movie and report the results of HR@10 in Fig.5(a). Smaller τ will lead to a decrease in sample matching while too large τ will fail to filter out potential unrelated item samples. Therefore we set $\tau = 0.5$. Finally we vary λ in $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1, 10\}$ on Amazon CD and Amazon Movie and report the results of HR@10 in Fig.5(b). The bell-shaped curve demonstrates a proper trade-off between intra-domain rating prediction and cross-domain private-robust embedding alignment and we set $\lambda = 0.1$.

5. Conclusion

In this paper, we propose Reducing Item Discrepancy (RidCDR) model on solving Privacy-Preserving Cross-Domain Recommendation (PPCDR) problem for inclusive public service. RidCDR involves rating prediction module and private-robust embedding alignment module for knowledge sharing privately. Specifically, we propose Differentially Private-Robust Adaptation (DPRA) method in private-robust embedding alignment module with two components, i.e., differentially private projection and robust reweighted sample adaptation. We reweight the data samples to filter out irrelevant items and utilize sample reweighted optimal transport for measuring the domain discrepancy. We also conduct extensive experiments to demonstrate the superior performance of our proposed RidCDR models.

Acknowledgement

This work was supported by National Key Rescarch and Development Program of China (No.2022YFF0902001), the National Research Foundation, Singapore and Infocomm Media Development Authority under its Trust Tech Funding Initiative.

Impact Statement

Previous privacy-preserving cross domain recommendation models should always rely on the overlapped users for knowledge sharing. In this paper, we consider a more general and challenging case when users and items are non-overlapped for privacy-preserving cross domain recommendation. The proposed Differentially Private-Robust Adaptation (DPRA) method in RidCDRcan significantly improve the results of trustworthy learning in recommendation scenarios for inclusive public service.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

An, D., Lei, N., Xu, X., and Gu, X. Efficient optimal transport algorithm by accelerated gradient descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 10119–10128, 2022.

Balaji, Y., Chellappa, R., and Feizi, S. Robust optimal transport with applications in generative modeling and domain adaptation. *Advances in Neural Information Processing Systems*, 33:12934–12944, 2020.

Benamou, J.-D. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(5):851–868, 2003.

Blocki, J., Blum, A., Datta, A., and Sheffet, O. The johnson-lindenstrauss transform itself preserves differential privacy. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 410–419. IEEE, 2012.

Blondel, M., Seguy, V., and Rolet, A. Smooth and sparse optimal transport. In *International conference on artificial intelligence and statistics*, pp. 880–889. PMLR, 2018.

Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Cao, J., Sheng, J., Cong, X., Liu, T., and Wang, B. Cross-domain recommendation to cold-start users via variational information bottleneck. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pp. 2209–2223. IEEE, 2022.

Cao, J., Li, S., Yu, B., Guo, X., Liu, T., and Wang, B. Towards universal cross-domain recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pp. 78–86, 2023.

Chai, D., Wang, L., Chen, K., and Yang, Q. Secure federated matrix factorization. *IEEE Intelligent Systems*, 36(5):11–20, 2020.

Chen, C., Zhang, M., Liu, Y., and Ma, S. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 world wide web conference*, pp. 1583–1592, 2018.

Chen, C., Wu, H., Su, J., Lyu, L., Zheng, X., and Wang, L. Differential private knowledge transfer for privacy-preserving cross-domain recommendation. In *Proceedings of the ACM Web Conference 2022*, pp. 1455–1465, 2022.

Chen, G., Zhang, X., Su, Y., Lai, Y., Xiang, J., Zhang, J., and Zheng, Y. Win-win: a privacy-preserving federated framework for dual-target cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4149–4156, 2023a.

Chen, X., Zhang, Y., Tsang, I. W., Pan, Y., and Su, J. Toward equivalent transformation of user preferences in cross domain recommendation. *ACM Transactions on Information Systems*, 41(1):1–31, 2023b.

Choi, Y., Choi, J., Ko, T., Byun, H., and Kim, C.-K. Based domain disentanglement without duplicate users or contexts for cross-domain recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 293–303, 2022.

de Oliveira, I. F. D. and Takahashi, R. H. C. Efficient solvers for armijo's backtracking problem. *arXiv preprint arXiv:2110.14072*, 2021.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Dwork, C. Differential privacy. In *International colloquium on automata, languages, and programming*, pp. 1–12. Springer, 2006.

Dwork, C. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pp. 1–19. Springer, 2008.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pp. 265–284. Springer, 2006.

Dwork, C., Roth, A., et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

Hao, X., Liu, Y., Xie, R., Ge, K., Tang, L., Zhang, X., and Lin, L. Adversarial feature translation for multi-domain recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 2964–2973, 2021.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pp. 173–182, 2017.

Hu, G., Zhang, Y., and Yang, Q. Conet: Collaborative cross networks for cross-domain recommendation. In *CIKM*, 2018.

Ju, H., Kang, S., Lee, D., Hwang, J., Jang, S., and Yu, H. Multi-domain recommendation to attract users via domain preference modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8582–8590, 2024.

Kang, S., Hwang, J., Lee, D., and Yu, H. Semi-supervised learning for cross-domain recommendation to cold-start users. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 1563–1572, 2019.

Kenthapadi, K., Korolova, A., Mironov, I., and Mishra, N. Privacy via the johnson-lindenstrauss transform. *arXiv preprint arXiv:1204.2606*, 2012.

Kim, D. and Fessler, J. A. Another look at the fast iterative shrinkage/thresholding algorithm (fista). *SIAM Journal on Optimization*, 28(1):223–250, 2018.

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

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Le, K., Nguyen, H., Nguyen, Q. M., Pham, T., Bui, H., and Ho, N. On robust optimal transport: Computational complexity and barycenter computation. *Advances in Neural Information Processing Systems*, 34:21947–21959, 2021.

Lê Tien, N., Habrard, A., and Sebban, M. Differentially private optimal transport: Application to domain adaptation. In *IJCAI*, pp. 2852–2858, 2019a.

Lê Tien, N., Habrard, A., and Sebban, M. Differentially private optimal transport: Application to domain adaptation. In *IJCAI*, pp. 2852–2858, 2019b.

Li, H., Ma, W., Sun, P., Li, J., Yin, C., He, Y., Xu, G., Zhang, M., and Ma, S. Aiming at the target: Filter collaborative information for cross-domain recommendation. *arXiv preprint arXiv:2403.20296*, 2024a.

Li, P. and Tuzhilin, A. Ddtcdr: Deep dual transfer cross domain recommendation. In *WSDM*, pp. 331–339, 2020.

Li, S., Yao, L., Mu, S., Zhao, W. X., Li, Y., Guo, T., Ding, B., and Wen, J.-R. Debiasing learning based cross-domain recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3190–3199, 2021.

Li, X., Qiu, Z., Zhao, X., Wang, Z., Zhang, Y., Xing, C., and Wu, X. Gromov-wasserstein guided representation learning for cross-domain recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 1199–1208, 2022.

Li, Y., Chen, C., Zhang, Y., Liu, W., Lyu, L., Zheng, X., Meng, D., and Wang, J. Ultrare: Enhancing receraser for recommendation unlearning via error decomposition. *Advances in Neural Information Processing Systems*, 36, 2024b.

Li, Y., Chen, C., Zheng, X., Liu, J., and Wang, J. Making recommender systems forget: Learning and unlearning for erasable recommendation. *Knowledge-Based Systems*, 283:111124, 2024c.

Li, Z., Amagata, D., Zhang, Y., Hara, T., Haruta, S., Yonekawa, K., and Kurokawa, M. Semantic relation transfer for non-overlapped cross-domain recommendations. In *PAKDD*, pp. 271–283. Springer, 2023.

Liao, X., Liu, W., Zheng, X., Yao, B., and Chen, C. Ppgencdr: A stable and robust framework for privacy-preserving cross-domain recommendation. *arXiv preprint arXiv:2305.16163*, 2023.

Liu, M., Li, J., Li, G., and Pan, P. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pp. 885–894, 2020.

Liu, W., Su, J., Chen, C., and Zheng, X. Leveraging distribution alignment via stein path for cross-domain cold-start recommendation. *Advances in Neural Information Processing Systems*, 34:19223–19234, 2021.

Liu, W., Zheng, X., Hu, M., and Chen, C. Collaborative filtering with attribution alignment for review-based non-overlapped cross domain recommendation. In *Proceedings of the ACM Web Conference 2022*, pp. 1181–1190, 2022a.

Liu, W., Zheng, X., Su, J., Hu, M., Tan, Y., and Chen, C. Exploiting variational domain-invariant user embedding for partially overlapped cross domain recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 312–321, 2022b.

Liu, W., Chen, C., Liao, X., Hu, M., Yin, J., Tan, Y., and Zheng, L. Federated probabilistic preference distribution modelling with compactness co-clustering for privacy-preserving multi-domain recommendation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 2206–2214, 2023a.

Liu, W., Zheng, X., Su, J., Zheng, L., Chen, C., and Hu, M. Contrastive proxy kernel stein path alignment for cross-domain cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2023b.

Liu, Y., Zhou, Z., and Sun, B. Cot: Unsupervised domain adaptation with clustering and optimal transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19998–20007, 2023c.

Lu, C., Yin, M., Shen, S., Ji, L., Liu, Q., and Yang, H. Deep unified representation for heterogeneous recommendation. *WWW*, 2022.

Man, T., Shen, H., Jin, X., and Cheng, X. Cross-domain recommendation: An embedding and mapping approach. In *IJCAI*, volume 17, pp. 2464–2470, 2017.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Meihan, W., Li, L., Tao, C., Rigall, E., Xiaodong, W., and Cheng-Zhong, X. Fedcdr: federated cross-domain recommendation for privacy-preserving rating prediction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 2179–2188, 2022.

Mukherjee, D., Guha, A., Solomon, J. M., Sun, Y., and Yurochkin, M. Outlier-robust optimal transport. In *International Conference on Machine Learning*, pp. 7850–7860. PMLR, 2021.

Ni, J., Li, J., and McAuley, J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *EMNLP*, pp. 188–197, 2019.

Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., and Talwar, K. Semi-supervised knowledge transfer for deep learning from private training data. *arXiv preprint arXiv:1610.05755*, 2016.

Pham, K., Le, K., Ho, N., Pham, T., and Bui, H. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, pp. 7673–7682. PMLR, 2020.

Pramod, D. Privacy-preserving techniques in recommender systems: state-of-the-art review and future research

agenda. *Data Technologies and Applications*, 57(1):32–55, 2023.

Qi, T., Wu, F., Wu, C., Huang, Y., and Xie, X. Privacy-preserving news recommendation model learning. *arXiv preprint arXiv:2003.09592*, 2020.

Qu, L., Yuan, W., Zheng, R., Cui, L., Shi, Y., and Yin, H. Towards personalized privacy: User-governed data contribution for federated recommendation. *arXiv preprint arXiv:2401.17630*, 2024.

Rakotomamonjy, A. and Liva, R. Differentially private sliced wasserstein distance. In *International Conference on Machine Learning*, pp. 8810–8820. PMLR, 2021.

Séjourné, T., Feydy, J., Vialard, F.-X., Trouvé, A., and Peyré, G. Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958*, 2019.

Séjourné, T., Vialard, F.-X., and Peyré, G. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pp. 4995–5021. PMLR, 2022.

Shu, R., Bui, H. H., Narui, H., and Ermon, S. A dirt-t approach to unsupervised domain adaptation. *ICLR*, 2018.

Van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Villani, C. et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Wang, C., Niepert, M., and Li, H. Recsys-dan: discriminative adversarial networks for cross-domain recommender systems. *IEEE transactions on neural networks and learning systems*, 31(8):2731–2740, 2019a.

Wang, X., He, X., Cao, Y., Liu, M., and Chua, T.-S. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–958, 2019b.

Wang, Y., Wang, Q., Zhao, L., and Wang, C. Differential privacy in deep learning: Privacy and beyond. *Future Generation Computer Systems*, 2023.

Wu, C., Wu, F., Lyu, L., Qi, T., Huang, Y., and Xie, X. A federated graph neural network framework for privacy-preserving personalization. *Nature Communications*, 13 (1):3091, 2022.

Xie, R., Liu, Q., Wang, L., Liu, S., Zhang, B., and Lin, L. Contrastive cross-domain recommendation in matching. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4226–4236, 2022.

Xie, Y., Wang, X., Wang, R., and Zha, H. A fast proximal point method for computing exact wasserstein distance. In *Uncertainty in artificial intelligence*, pp. 433–453. PMLR, 2020.

Xu, K., Li, C., Zhu, J., and Zhang, B. Understanding and stabilizing gans' training dynamics using control theory. In *International Conference on Machine Learning*, pp. 10566–10575. PMLR, 2020.

Yan, D., Zhao, Y., Yang, Z., Jin, Y., and Zhang, Y. Fedcdr: Privacy-preserving federated cross-domain recommendation. *Digital Communications and Networks*, 8(4): 552–560, 2022.

Yang, M., Guo, T., Zhu, T., Tjuawinata, I., Zhao, J., and Lam, K.-Y. Local differential privacy and its applications: A comprehensive survey. *Computer Standards & Interfaces*, pp. 103827, 2023a.

Yang, W., Yang, J., and Liu, Y. Multimodal optimal transport knowledge distillation for cross-domain recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 2959–2968, 2023b.

Yu, W., Lin, X., Ge, J., Ou, W., and Qin, Z. Semi-supervised collaborative filtering by text-enhanced domain adaptation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2136–2144, 2020.

Zang, T., Zhu, Y., Liu, H., Zhang, R., and Yu, J. A survey on cross-domain recommendation: taxonomies, methods, and future directions. *TOIS*, 41(2):1–39, 2022.

Zhang, Q., Liao, W., Zhang, G., Yuan, B., and Lu, J. A deep dual adversarial network for cross-domain recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2021a.

Zhang, T., Chen, C., Wang, D., GuoMember, J., and Song, B. A vae-based user preference learning and transfer framework for cross-domain recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

Zhang, Y., Liu, Y., Han, P., Miao, C., Cui, L., Li, B., and Tang, H. Learning personalized itemset mapping for cross-domain recommendation. In *IJCAI*, pp. 2561–2567, 2021b.

Zhao, C., Li, C., and Fu, C. Cross-domain recommendation via preference propagation graphnet. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2165–2168, 2019.

Zhao, C., Li, C., Xiao, R., Deng, H., and Sun, A. Catn: Cross-domain recommendation for cold-start users via

aspect transfer network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 229–238, 2020a.

Zhao, C., Zhao, H., He, M., Zhang, J., and Fan, J. Cross-domain recommendation via user interest alignment. In *Proceedings of the ACM Web Conference 2023*, pp. 887–896, 2023.

Zhao, Y., Zhao, J., Yang, M., Wang, T., Wang, N., Lyu, L., Niyato, D., and Lam, K.-Y. Local differential privacy-based federated learning for internet of things. *IEEE Internet of Things Journal*, 8(11):8836–8853, 2020b.

Zheng, L., Noroozi, V., and Yu, P. S. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth ACM international conference on web search and data mining*, pp. 425–434, 2017.

Zhu, F., Chen, C., Wang, Y., Liu, G., and Zheng, X. Dtcdr: A framework for dual-target cross-domain recommendation. In *CIKM*, pp. 1533–1542, 2019.

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

Zhu, F., Wang, Y., Zhou, J., Chen, C., Li, L., and Liu, G. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 2021b.

Zhu, J., Wang, Y., Zhu, F., and Sun, Z. Domain disentanglement with interpolative data augmentation for dual-target cross-domain recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pp. 515–527, 2023.

Zhu, Y., Yu, X., Tsai, Y.-H., Pittaluga, F., Faraki, M., Wang, Y.-X., et al. Voting-based approaches for differentially private federated learning. *arXiv preprint arXiv:2010.04851*, 2020.

Zhu, Y., Tang, Z., Liu, Y., Zhuang, F., Xie, R., Zhang, X., Lin, L., and He, Q. Personalized transfer of user preferences for cross-domain recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 1507–1515, 2022.

Appendix

A. Dual Form of UOT

To start with, we first illustrate the definition of UOT. Unlike balanced optimal transport, which imposes strict mass equality constraints across domains, UOT relaxes these constraints while improving feasibility in real-world applications (Benamou, 2003; Séjourné et al., 2022). Specifically, the formulation of UOT is given as:

$$\min_{\boldsymbol{\pi}\geq 0} J_{\mathrm{UOT}} = \left[\langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau \mathrm{KL}(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}) + \tau \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_N \| \boldsymbol{b})\right],$$

where $\boldsymbol{C} \in \mathbb{R}^{N\times N}$ denotes the cost matrix and it can be calculated via $C_{ij} = ||\widehat{\boldsymbol{V}}_i^{(\mathrm{src})} - \widehat{\boldsymbol{V}}_j^{(\mathrm{trg})}||_2^2$. $\boldsymbol{a}$, $\boldsymbol{b}$ denote the initial sample weights and we set them $a_i = b_j = \frac{1}{N}$ respectively. $\mathrm{KL}(\boldsymbol{x}||\boldsymbol{y})$ denotes the KL Divergence between two d-dimensional data samples $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$ as $\mathrm{KL}(\boldsymbol{x}||\boldsymbol{y}) = \sum_{i=1}^d \left[x_i \log \frac{x_i}{y_i} - x_i + y_i\right]$. Here τ denotes the balanced hyper parameter.

Then we deduce the Fenchel-Lagrange conjugate form of UOT.

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \tau \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) \\ &s.t.\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}, \end{aligned} \tag{9}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the marginal probabilities for source and target domains respectively. The Lagrange multipliers of UOT with KL-Divergence is given as:

$$\begin{aligned} \max_{\boldsymbol{u},\boldsymbol{v},\boldsymbol{s},\zeta} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J} &= \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \tau \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi}\mathbf{1}_N - \boldsymbol{\alpha}\rangle + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top \mathbf{1}_M - \boldsymbol{\beta}\rangle - \langle \boldsymbol{s}, \boldsymbol{\pi}\rangle \\ &= \tau \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle + \tau \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top \mathbf{1}_M\rangle + \sum_{i,j} (C_{ij} - u_i - v_j - s_{ij})\pi_{ij}, \end{aligned} \tag{10}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$ and ζ denote as the multipliers. Note that $s_{ij}\pi_{ij} = 0$ and $s_{ij} \geq 0$. Meanwhile we should notice that (using $\mathrm{KL}(\boldsymbol{\pi}\mathbf{1}_N||\boldsymbol{a})$ as an example):

$$\begin{aligned} \frac{\partial}{\partial \pi_{ij}} \left(\mathrm{KL}(\boldsymbol{\pi}\mathbf{1}_N||\boldsymbol{a})\right) &= \frac{\partial}{\partial \pi_{ij}} \left(\sum_{i=1}^M \left[\sum_{j=1}^N \pi_{ij} \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} - \sum_{j=1}^N \pi_{ij} + a_i \right] \right) \\ &= \frac{\partial}{\partial \sum_{j=1}^N \pi_{ij}} \left(\sum_{i=1}^M \left[\frac{\sum_{j=1}^N \pi_{ij}}{a_i} \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} - \frac{\sum_{j=1}^N \pi_{ij}}{a_i} + 1 \right] a_i \right) \frac{\partial \sum_{j=1}^N \pi_{ij}}{\partial \pi_{ij}} \\ &= \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i}. \end{aligned} \tag{11}$$

By taking the differentiation on π_{ij}:

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial \pi_{ij}} &= \left[\tau \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} + u_i + \zeta\right] + \left[\tau \log \frac{\sum_{i=1}^N \pi_{ij}}{b_j} + v_j - \zeta\right] + (C_{ij} - u_i - v_j - s_{ij}) \\ &= C_{ij} + \tau \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} + \tau \log \frac{\sum_{i=1}^N \pi_{ij}}{b_j} - s_{ij} \\ &= 0. \end{aligned} \tag{12}$$

We can observe that:

$$\sum_{j=1}^{N}\pi_{ij}=a_i\exp\left(-\frac{u_i+\zeta}{\tau_a}\right)\quad\text{and}\quad\sum_{i=1}^{N}\pi_{ij}=b_j\exp\left(-\frac{v_j-\zeta}{\tau_b}\right)\quad\text{and}\quad C_{ij}-u_i-v_j-s_{ij}=0. \tag{13}$$

Then we take them back into KL-Divergence as (using $\tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}\right)+\langle\boldsymbol{u}+\zeta,\boldsymbol{\pi}\mathbf{1}_N\rangle$ as an example):

$$\begin{aligned}
&\tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}\right)+\langle\boldsymbol{u}+\zeta,\boldsymbol{\pi}\mathbf{1}_N\rangle\\
=&\tau\mathrm{KL}\left(\left\langle\boldsymbol{a},\exp\left(-\frac{\boldsymbol{u}+\zeta}{\tau}\right)\right\rangle\|\boldsymbol{a}\right)+\left\langle\boldsymbol{u}+\zeta,\boldsymbol{a}\exp\left(-\frac{\boldsymbol{u}+\zeta}{\tau}\right)\right\rangle\\
=&\tau\sum_{i=1}^{N}\left[a_i\exp\left(-\frac{u_i+\zeta}{\tau}\right)\log\frac{a_i\exp\left(-\frac{u_i+\zeta}{\tau}\right)}{a_i}-a_i\exp\left(-\frac{u_i+\zeta}{\tau}\right)+a_i\right]+\sum_{i=1}^{N}(u_i+\zeta)a_i\exp\left(-\frac{u_i+\zeta}{\tau}\right)\\
=&\sum_{i=1}^{N}\left[-\tau a_i\exp\left(-\frac{u_i+\zeta}{\tau}\right)+\tau a_i\right].
\end{aligned} \tag{14}$$

Therefore, we can obtain the Fenchel-Lagrange conjugate form of UOT with variables ζ, $\boldsymbol{u}$ and $\boldsymbol{v}$ as:

$$\begin{aligned}
&\min_{\boldsymbol{v},\boldsymbol{u},\zeta}L_{\mathrm{UOT}}=\tau\left[e^{-\frac{\zeta}{\tau}}\langle\boldsymbol{a},e^{-\frac{\boldsymbol{u}}{\tau}}\rangle+e^{\frac{\zeta}{\tau}}\langle\boldsymbol{b},e^{-\frac{\boldsymbol{v}}{\tau}}\rangle\right]\\
&s.t.\ u_i+v_j\le C_{ij}.
\end{aligned} \tag{15}$$

Algorithm 3 The optimization procedure on ℓ_{RDKF} via FISTA

Input: ℓ_{RDKF}: The function should be optimized.
Procedure:

1: Initialize $t=0$, $\boldsymbol{f}^{(0)}=(0,0,\cdots,0)$ and $\theta_0=1$.
2: **repeat**
3: Calculate the sub-gradient $\mathcal{G}(f^{(t)})$ as:

$$\mathcal{G}(f_i^{(t)})=-\widehat{a}_i+\sum_{j=1}^{N}\delta_i\left(\sup_{k\in[N]}\left(f_k^{(t)}-C_{kj}\right)\right)\cdot\widehat{b}_j\quad\text{where}\quad\delta_i\left(\sup_{k\in[N]}\left(f_k^{(t)}-C_{kj}\right)\right)=\begin{cases}1,k^*=i\\0,\text{Others}\end{cases} \tag{16}$$

.
4: Find Step via $\eta_t=\mathrm{LineSearch}(\boldsymbol{f}^{(t)},\ell_{\mathrm{RDKF}}(\boldsymbol{f}^{(t)}),\mathcal{G}(\boldsymbol{f}^{(t)}))$.
5: Update $\widehat{\boldsymbol{f}}^{(t+1)}=\boldsymbol{f}^{(t)}-\eta_t\mathcal{G}(\boldsymbol{f}^{(t)})$.
6: Update $\theta_{t+1}=\frac{1}{2}(1+\sqrt{1+4\theta_t^2})$
7: Update $\boldsymbol{f}^{(t+1)}=\boldsymbol{f}^{(t)}+\frac{\theta_t-1}{\theta_{t+1}}(\widehat{\boldsymbol{f}}^{(t+1)}-\boldsymbol{f}^{(t)})$.
8: Make projection $\boldsymbol{f}^{(t+1)}\leftarrow\left[\boldsymbol{f}^{(t+1)}-\mathrm{mean}(\boldsymbol{f}^{(t+1)})\right]$.
9: Update $t=t+1$.
10: **until** Converge
11: **Return**: The optimal $\boldsymbol{f}^o=\boldsymbol{f}^{(T)}$ after T-th iteration.

B. FISTA Algorithm of SROT

The *Reweighting Discrete Kantorovich Functional (RDKF)* as given as below:

$$\max_{\boldsymbol{f}\in\boldsymbol{\Delta}}\ell_{\mathrm{RDKF}}=\sum_{i=1}^{N}f_i\widehat{a}_i-\sum_{j=1}^{N}\widehat{b}_j\sup_{k\in[N]}(f_k-C_{kj}) \tag{17}$$

Note that we only need to optimize $\boldsymbol{f}$ during the whole optimization process. We further add zero-mean constraints on $\boldsymbol{f}$ as $\Delta=\{\sum_{i=1}^{N}f_i=0\}$. Likewise, we further adopt FISTA algorithm to optimize Eq.(17) for finding the optimal solution on $\boldsymbol{f}^*$. The optimization details can be found in Algo.3.

Datasets	Users	Items	Ratings	Density
Douban Movie	20,464	18,173	446,821	0.12%
Douban Book	18,867	17,019	311,250	0.09%
Douban Music	16,230	13,647	224,541	0.10%
Amazon Movie	19,736	18,224	124,383	0.034%
Amazon Book	20,569	19,827	131,945	0.032%
Amazon CDs	15,041	11,955	46,348	0.025%
Amazon Videos	9,625	8,568	11,676	0.014%
Amazon Games	13,222	10,080	13,377	0.010%

Table 3. Statistics of Douban and Amazon datasets.

C. Datasets

Datasets. We conduct extensive experiments on two popuarly used real-world datasets, i.e., *Amazon* and *Douban*. The **Amazon** dataset (Ni et al., 2019) has five datasets, i.e., Movie, Book, CD, Video, and Game. The **Douban** dataset (Zhu et al., 2019; 2021b) has three domains, i.e., Book, Music, and Movie. The detailed statistics of these datasets are given in Table.3. For each dataset, we binarize the ratings higher or equal to 4 as positive. We also filter the users and items with less than 5 interactions, following existing research (Zhu et al., 2019; Liu et al., 2022b).