CausalCDR: Causal Embedding Learning for Cross-domain Recommendation

Fengxin Li * Hongyan Liu † Jun He ‡ Xiaoyong Du §

Abstract

Cross-domain recommendation (CDR) methods achieve success in disentangling user preferences into domain-specific and domain-shared parts. However, recent research has shown that isolated domain-specific preference limits performance improvements. In this paper, we propose a new CDR framework, called CausalCDR, which identifies the limitations of existing methods and addresses existing issues. CausalCDR consists of two views: the causal view and the generative view. The causal view incorporates causality of variables into the CDR scenario, while the generative view implements the causal view by modeling the joint distribution of user interaction via encoding, causal, and generation stage. To optimize CausalCDR, we re-derive the Evidence Lower Bound (ELBO) and introduce a mutual information regularizer and an adversarial classifier. We evaluate CausalCDR on four real-world CDR scenarios and demonstrate its effectiveness in improving CDR performance. **Keywords:** Cross-domain recommendation, Causal Inference, Variational Inference

1 Introduction

Recommendation systems have been applied in various practical application areas such as entertainment and e-commerce [23, 43]. However, conventional recommendation methods suffer from issues such as cold start and data sparsity [7, 28]. As a result, cross-domain recommendation (CDR) has gained significant attention from both academia and industry [5, 52].

Motivated by disentangled embedding learning [8, 47], recent studies [3, 41] have attempted to disentangle domain-specific and domain-shared embedding from the user embedding. By identifying common preferences across domains, sharing embedding links the different domains and contributes to performance increase. However, while disentanglement methods have shown outstanding performance, they typically focus only on the domain-shared preference among domains, which leaves domain-specific preferences isolated between domains.

Figure 1(a) illustrates an example for the Phone&Clothing scenario. In this scenario, there are domain-shared preferences for features like "Price" in both domains, as well as domain-specific preferences for features such as "Processor" and "Battery" for Phone and "Appearance" and "Materials" for Clothing. We have discovered that user groups exhibit similar behavior not only in domain-shared preferences but also in domain-specific preferences [9, 22]. For instance, a younger consumer group tends to purchase phones with moderate price, high-performance processors and normal battery capacity, as well as clothes with moderate price, stylish appearance and arbitrary materials. Similarly, a more mature consumer group tends to purchase phones with lower prices, normal processors, and large battery capacity, as well as clothes with lower prices, stylish appearance, and firm materials. Both the domain-shared preference and domain-specific preference have the implicit link between domains. Unfortunately, existing methods fail to recognize the link between domain-specific preferences.

To address the above issue, we propose a novel CDR framework called CausalCDR (Causal embedding learning for Cross-Domain Recommendation). In CausalCDR, we build modes from two views: causal view and generative view. The causal view incorporates causality modeling into the CDR scenario for domain x and y. As illustrated in Figure 1(b), user preference is disentangled into domain-shared, U^S, and domain-specific preference, U^x, U^y. To tackle the isolated preference issue, we model a common cause of domain-shared and domain-specific preference, the user profile information Z. Specifically, the domain-shared preference represents the common preference across domains that are only influenced by individual user profile. Therefore, no matter how the domain changes, the domain-shared preference will remain consistent. On the other hand, the domain-specific preference is influenced by both the user profile and the domain information denoted by D^x or D^y, changing as the domain shifts.

*lifengxin@ruc.edu.cn, Renmin University of China.

†hyliu@tsinghua.edu.cn, Tinghua University, corresponding author.

‡hejun@ruc.edu.cn, Renmin University of China.

§duyong@ruc.edu.cn, Renmin University of China.

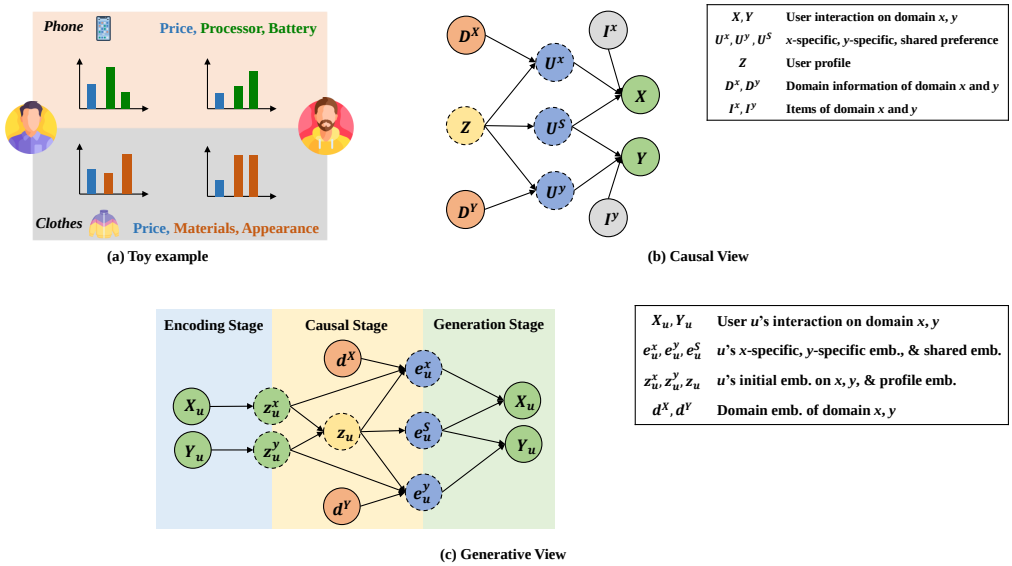


Figure 1: (a): An example of domain-shared preference and domain-specific preference. (b): Causal View of proposed CausalCDR. (c): Generative View of proposed CausalCDR

To learning embeddings for variables modeled in the causal view, we build model from a generative view illustrated in Figure 1(c) following VAE framework [16, 46]. The generative view models the joint distribution of user interaction on two domains through three stages: encoding stage, causal stage and generation stage. The encoding stage estimates the posterior distribution to obtain the initial hidden embeddings z_u^x and z_u^y of user interaction X_u and Y_u. The causal stage decomposes the posterior according to the causal view and conducts causal inference to infer the user causal embeddings e_u^x, e_u^y and e_u^S. The generation stage estimates the conditional joint distribution to generate user interaction X_u and Y_u via the causal embeddings.

To further optimize the proposed method, CausalCDR re-derives the Evidence Lower Bound (ELBO) and introduces a mutual information regularizer and an adversarial domain classifier. The regularizer aims to maximize the mutual information between profile embedding and initial hidden embedding to facilitate the learning of user profile. The domain classifier encourages the disentanglement of domain-specific and domain-shared features of items.

We conduct extensive experiments on four real-world CDR scenarios, demonstrating that CausalCDR outperforms existing CDR methods. Additionally, ablation studies validate the effectiveness of CausalCDR's components.

Our contributions can be summarized as follows: (1) We investigate existing CDR methods and identify the problem of isolated specific preference. (2) To tackle this issue, we propose a new CDR framework called CausalCDR, consisting of causal view and generative view. The causal view model incorporates causality into the CDR scenario, while the generative view implements causal view by modeling the joint distribution of user interaction. (3) Furthermore, we re-derive the Evidence Lower Bound (ELBO), and propose a mutual information regularizer and an adversarial domain classifier to optimize our proposed method. (4) Extensive experiments on four real-world CDR scenarios demonstrate that CausalCDR outperforms existing CDR methods.

2 CausalCDR

In this section, we provide a detailed description of CausalCDR from both the causal and generative views. The causal view identifies causality in the CDR scenario, while the generative view models the joint distribution of interactions to implement the causal view. To optimize the proposed method, we re-derive the Evidence Lower Bound (ELBO), and introduce a mutual information regularizer and an adversarial domain classifier.

2.1 Causal View In the causal view, we model the interaction generation process and formulate CDR through a causal graph illustrated in Figure 1(b). The dotted circle indicates unobserved variables. The meanings of nodes and edges are detailed as follows:

- X and Y represent user interactions on domains x and y, respectively.
- U^x, U^y, and U^S represent x-specific preference (e.g., preference for Processor, Battery in Phone do-

main), y-specific preference (e.g., preference for Appearance, Materials in Clothing domain), and domain-shared preference (e.g., preference for Price in both Phone and Clothing domain) respectively.

• Z represents the general user profile information(e.g., age, gender).

• D^x and D^y represent domain information for domains x and y respectively.

• I^x and I^y represent items for domains x and y respectively.

• $(U^x, U^S, I^x) \to X$ and $(U^y, U^S, I^y) \to Y$ denote that user interaction on each domain (x or y) is determined by both domain-specific and domain-shared preference, as well as items of each domain.

• $(D^x, Z) \to U^x$ and $(D^y, Z) \to U^y$ indicate that domain-specific preference is determined by both user profile and domain information, leading to changes following domain shifts.

• $Z \to U^S$ indicates that domain-shared preference is solely determined by the user profile. There is no causal link between domain-shared preference and domain information. Therefore, no matter how the domain changes, the domain-shared preference will remain consistent.

Overall, the causal view provides a comprehensive understanding of the interaction generation process and the causal relationships among user preference, domain information, and user profile in CDR.

2.2 Generative View Due to the unobserved variables, it is difficult to infer the causality directly through the observed variables. Therefore, we adopt the principles of variational inference [16, 46] to build a generative view for implementation. In generative view, we formulate the CDR problem, derive a tractable log-likelihood, and implement the probabilistic estimators through neural networks.

2.2.1 Problem Formulation for CDR Let $\mathcal{U}^x$ be the set of users in domain x and $\mathcal{U}^y$ be the set of users in domain y. We assume that $\mathcal{U}^x = \mathcal{U}^y = \mathcal{U}$, which means that every user u in the user set $\mathcal{U}$ has interactions on both domains x and y. Let $\mathcal{I}^x$ be the set of items in domain x and $\mathcal{I}^y$ be the set of items in domain y. We assume that $\mathcal{I}^x \cap \mathcal{I}^y = \varnothing$ to satisfy the cross-domain scenario. The interaction vectors of user u are denoted by $X_u \in \mathbb{R}^{|\mathcal{I}^x|}$ and $Y_u \in \mathbb{R}^{|\mathcal{I}^y|}$. Thus, the dataset $\mathcal{O}$ is composed of $|\mathcal{U}|$ users of the form (u, X_u, Y_u).

2.2.2 Generative View As illustrated in Figure 1(c), the generation view of the model composes of three stages:

Encoding stage: Encoding the interaction vectors X_u and Y_u to obtain their initial embeddings $z_u^x \in \mathbb{R}^m$ and $z_u^y \in \mathbb{R}^m$ respectively.

Causal stage: Conduct causal inference to obtain the disentangled causal embeddings e_u^x, e_u^y, and e_u^S. According to the causal graph shown in Figure 1(b), we first disentangle both domains' interaction embeddings z_u^x and z_u^y to user's profile embedding z_u. User's domain-share preference $e_u^S \in \mathbb{R}^{m/2}$ is inferred directly from z_u, where m denotes the vector size, as it is independent of domain information. Domain-x-specific preference $e_u^x \in \mathbb{R}^{m/2}$ is inferred based on z_u^x, z_u, and domain embedding $d^x \in \mathbb{R}^m$. Similarly, domain-y-specific preference $e_u^y \in \mathbb{R}^{m/2}$ is inferred based on z_u^y, z_u, and domain embedding $d^y \in \mathbb{R}^m$.

Generation stage: Reconstruct the interaction X_u and Y_u using causal embeddings e_u^x, e_u^y, and e_u^S.

To construct the generative view, we first provide a probabilistic formulation for the joint distribution $p(X_u, Y_u | d^x, d^y)$. For a sample (u, X_u, Y_u) from the dataset $\mathcal{O}$, we consider the initial embeddings z_u^x, z_u^y, the profile embedding z_u and the causal embeddings e_u^x, e_u^S, e_u^y as latent variables, with the domain information d^x and d^y as conditions. Specifically, we consider the following conditional generative model:

$$\begin{aligned} & p(X_u, Y_u | d^x, d^y) \\ = & \int p(X_u, Y_u | d^x, d^y, z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y) \\ & dz_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y \end{aligned} \tag{2.1}$$

To derive the tractable approximation of Equation 2.1, we introduce a variational distribution $q(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | X_u, Y_u, d^x, d^y)$. The generative view solves for both the joint distribution and variational distribution. In generation stage, the joint distribution is decomposed into:

$$\begin{aligned} & p(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y, X_u, Y_u | d^x, d^y) \\ = & p_{\theta^x}(X_u | e_u^x, e_u^S) p_{\theta^y}(Y_u | e_u^y, e_u^S) \\ & p(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | d^x, d^y) \end{aligned} \tag{2.2}$$

where θ^x, θ^y denote the parameters of decoders respectively. After decomposing, the joint distribution is divided two parts: decoders $p_{\theta^x}, p_{\theta^y}$ and prior distribution $p(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | d^x, d^y)$. We define the prior distribution based on the causal view:

$$\begin{aligned} & p(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | d^x, d^y) \\ = & p(e_u^x | z_u^x, z_u, d^x) p(e_u^y | z_u^y, z_u, d^y) p(e_u^S | z^u) \\ & p(z_u | z_u^x, z_u^y) p(z_u^x | d^x) p(z_u^y | d^y) \end{aligned} \tag{2.3}$$

In the encoding stage, the variational distribution is decomposed into:

$$\begin{aligned} & q(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | X_u, Y_u, d^x, d^y) \\ = & q_{\psi^x}(z_u^x | X_u) q_{\psi^y}(z_u^y | Y_u) \\ & q_{\psi^C}(z_u, e_u^x, e_u^S, e_u^y | z_u^x, z_u^y, d^x, d^y) \end{aligned} \tag{2.4}$$

where ψ^x, ψ^y are the parameters of encoders and ψ^C denotes the parameter of causal stage. Similarly, after

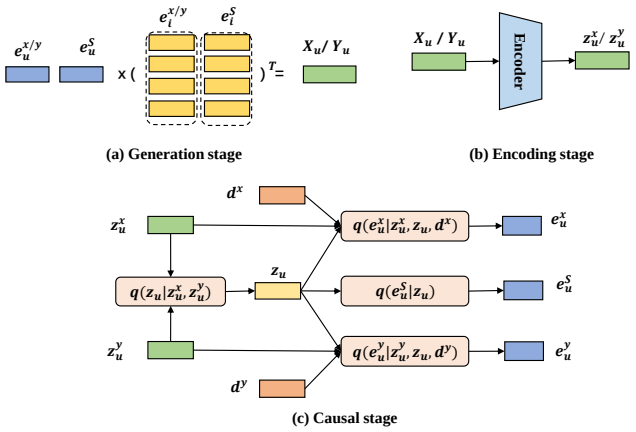


Figure 2: Implementation: (a): Generation Stage. (b): Encoding Stage. (c): Causal Stage.

decomposition, the posterior distribution is divided into two parts: encoders q_{ψ^x}, q_{ψ^y}, and posterior distribution $q_{\psi^C}(z_u, e_u^x, e_u^S, e_u^y | z_u^x, z_u^y, d^x, d^y)$.

To bridge the gap between the generation stage and encoding stage and to generate disentangled causal embeddings, we designed the causal stage based on the causal relationship of variables. As shown in Figure 1(c), we assume that there is no relationship between nodes without an edge. Thus, the posterior distribution can be decomposed into:

$$\begin{aligned} &q_{\psi^C}(z_u, e_u^x, e_u^S, e_u^y | z_u^x, z_u^y, d^x, d^y) \\ =&q_{\psi_x^C}(e_u^x | z_u^x, z_u, d^x) q_{\psi_y^C}(e_u^y | z_u^y, z_u, d^y) \\ &q_{\psi_S^C}(e_u^S | z_u) q_{\psi_z^C}(z_u | z_u^x, z_u^y) \end{aligned} \tag{2.5}$$

where $\psi^C = \{\psi_z^C, \psi_x^C, \psi_y^C, \psi_S^C\}$.

2.2.3 Implementation After decomposing the joint distribution and variational distribution, we give the implementation of the proposed generative view.

Generation Stage. The interaction vectors X_u and Y_u are assumed to follow multinomial priors [33]. Specifically, X_u and Y_u are modeled as:

$$\begin{aligned} X_u &\sim Mult(|\mathcal{I}^x|, \Phi(f_{\theta^x}(e_u^x, e_u^S))), \\ Y_u &\sim Mult(|\mathcal{I}^y|, \Phi(f_{\theta^y}(e_u^y, e_u^S))) \end{aligned} \tag{2.6}$$

Here, $\Phi(\cdot)$ denotes the *sigmoid* function to normalize the output f_θ. We implement the function f_{θ^x} and f_{θ^y} using Matrix Factorization [13]. As illustrated in Figure 2(a), the interaction vectors are generated as:

$$\begin{aligned} X_u &= (e_u^x || e_u^S) \cdot (e_{i^{x:}}^x || e_{i^{y:}}^S)^T, \\ Y_u &= (e_u^y || e_u^S) \cdot (e_{i^{y:}}^y || e_{i^{y:}}^S)^T \end{aligned} \tag{2.7}$$

Here, $(\cdot||\cdot)$ denotes the concatenation operation. $i^{x:}, i^{y:}$ denotes all the items $i^x \in \mathcal{I}^x$ and $i^y \in \mathcal{I}^y$, and $e_{i^x}^x \in \mathbb{R}^{m/2}$, $e_{i^x}^S \in \mathbb{R}^{m/2}$, $e_{i^y}^y \in \mathbb{R}^{m/2}$, $e_{i^y}^S \in \mathbb{R}^{m/2}$ denote the domain-specific feature and domain-shared feature embedding of item i^x and i^y.

Encoding Stage. The initial embedding z_u^x and z_u^y are assumed to follow normal priors [16]. Specifically, z_u^x and z_u^y are modeled as:

$$\begin{aligned} z_u^x &\sim \mathcal{N}(\mu_{\psi^x}(X_u), \text{diag}\{\sigma_{\psi^x}^2(X_u)\}), \\ z_u^y &\sim \mathcal{N}(\mu_{\psi^y}(Y_u), \text{diag}\{\sigma_{\psi^y}^2(Y_u)\}) \end{aligned} \tag{2.8}$$

Here $\text{diag}\{\cdot\}$ denotes the diagonal covariance of the normal distribution. The function $\mu_{\psi^x}(\cdot)$, $\sigma_{\psi^x}^2(\cdot)$, $\mu_{\psi^y}(\cdot)$, and $\sigma_{\psi^y}^2(\cdot)$ are implemented by multi-layer neural networks. The reparameterization trick is utilized to enable gradient backpropagation. Additionally, in this stage, we extract high-dimensional domain embeddings by:

$$\begin{aligned} d^x &= g_\omega(D^x || \text{mean-pooling}(e_{i^{x:}}^x)), \\ d^y &= g_\omega(D^y || \text{mean-pooling}(e_{i^{y:}}^y)) \end{aligned} \tag{2.9}$$

Here $D^x \in \mathbb{R}^{m/2}$ and $D^y \in \mathbb{R}^{m/2}$ denote learnable domain embeddings respectively, and g_ω denotes the extractor implemented by multi-layer neural network.

Causal Stage. Similar to the Encoding stage, we assume that the profile embedding z_u and causal embedding e_u^x, e_u^S, e_u^y follows normal priors as:

$$\begin{aligned} e_u^x &\sim \mathcal{N}(\mu_{\psi_x^C}(z_u, z_u^x, d^x), \text{diag}\{\sigma_{\psi_x^C}^2(z_u, z_u^x, d^x)\}), \\ e_u^y &\sim \mathcal{N}(\mu_{\psi_y^C}(z_u, z_u^y, d^x), \text{diag}\{\sigma_{\psi_y^C}^2(z_u, z_u^y, d^x)\}), \\ e_u^S &\sim \mathcal{N}(\mu_{\psi_S^C}(z_u), \text{diag}\{\sigma_{\psi_S^C}^2(z_u)\}), \\ z_u &\sim \mathcal{N}(\mu_{\psi_z^C}(z_u^x, z_u^y), \text{diag}\{\sigma_{\psi_z^C}^2(z_u^x, z_u^y)\}) \end{aligned} \tag{2.10}$$

The function $\mu_{\psi_z^C}(\cdot)$, $\sigma_{\psi_z^C}^2(\cdot)$, $\mu_{\psi_x^C}(\cdot)$, $\sigma_{\psi_x^C}^2(\cdot)$, $\mu_{\psi_y^C}(\cdot)$, $\sigma_{\psi_y^C}^2(\cdot)$, $\mu_{\psi_S^C}(\cdot)$, and $\sigma_{\psi_S^C}^2(\cdot)$ are implemented by multi-layer neural networks.

2.3 Training Objective In this section, we discuss the training objective of CausalCDR, which aims to learn the causal embedding. Our training objective considers three main concerns: (1) Evidence Lower BOund (ELBO) of generative view. (2) Mutual information regularizer for profile embedding. (3) Adversarial domain classifier for disentangled causal embedding.

ELBO of generative view: Given the mini batch $\mathcal{B}$ of dataset $\mathcal{O}$, We apply variational inference to solve the log-likelihood of the joint distribution $p(X_u, Y_u | d^x, d^y)$.

$$\begin{aligned} &\mathbb{E}_{q_\mathcal{B}}[\log p(X_u, Y_u | d^x, d^y)] \leq -L_{ELBO} \\ =&\mathbb{E}_{q_\mathcal{B}}[\mathbb{E}_{q(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | X_u, Y_u, d^x, d^y)}[\\ &\log \frac{p(X_u, Y_u | d^x, d^y, z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y)}{q(z_u^x, z_u^y, z_u, e_u^x, e_u^S, e_u^y | X_u, Y_u, d^x, d^y)}]] \end{aligned} \tag{2.11}$$

According to the decomposition of Equation 2.2-2.5, Equa-

tion 2.11 could be decomposed into:

$$\begin{aligned}
L_{ELBO} = & - \mathbb{E}_{q_{\mathcal{B}}}[\mathbb{E}_{q_\psi(e_u^x, e_u^S | X_u, Y_u, d^x, d^y)}[\log p_{\theta^x}(X|e^x, e^S, d^x)]] \\
& - \mathbb{E}_{q_{\mathcal{B}}}[\mathbb{E}_{q_\psi(e_u^S, e_u^y | X_u, Y_u, d^x, d^y)}[\log p_{\theta^y}(Y|e^y, e^S, d^y)]] \\
& + \mathbb{D}_{KL}(q_\psi(e_u^x|z_u^x, z_u, d^x)||p(e_u^x|z_u^x, z_u, d^x)) \\
& + \mathbb{D}_{KL}(q_\psi(e_u^y|z_u^y, z_u, d^y)||p(e_u^y|z_u^y, z_u, d^y)) \\
& + \mathbb{D}_{KL}(q_\psi(e_u^S|z_u)||p(e_u^S|z_u)||) \\
& + \mathbb{D}_{KL}(q_\psi(z_u|z_u^x, z_u^y)||p(z_u|z_u^x, z_u^y)) \\
& + \mathbb{D}_{KL}(q_\psi(z_u^x|X_u)||p(z_u^x|d^x)) \\
& + \mathbb{D}_{KL}(q_\psi(z_u^y|Y_u)||p(z_u^y|d^y))
\end{aligned}$$

Here $\mathbb{D}_{KL}(\cdot||\cdot)$ denotes the KL-divergence. The first two terms in L_{ELBO} are implemented as the reconstruction loss, denoted by L_{re}, while the other terms are the KL-divergences between the prior and posterior, denoted by L_{KL}. Following previous works [16, 33], we assume that all priors follow a standard normal distribution, $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Mutual information regularizer for profile embedding: The user profile embedding z_u captures user characteristics that determines user's both domain-specific and domain-shared preference. We assume that z_u is the common between domain initial embeddings z_u^x and z_u^y. To achieve this, we use mutual information to measure the information shared between z_u and the initial hidden embeddings z_u^x and z_u^y. The objective is defined as: $\max \text{MI}(z_u, z_u^x) + \text{MI}(z_u, z_u^y)$, where $\text{MI}(\cdot, \cdot)$ denotes mutual information. To estimate the mutual information, we use InfoNCE [32], which is a contrastive learning approach that has been shown to work well for high-dimensional embeddings. Specifically, we define L_{MI} loss as:

$$L_{MI} = -\log \frac{\exp(z_u(z_u^x)^T)}{\Sigma_{j \in \mathcal{B}} \exp(z_u(z_j^x)^T)} - \log \frac{\exp(z_u(z_u^y)^T)}{\Sigma_{j \in \mathcal{B}} \exp(z_u(z_j^y)^T)}$$

where j denotes user j in the mini-batch $\mathcal{B}$. L_{MI} aims to maximize the dot product between z_u and its corresponding initial embedding z_u^x (or z_u^y) at the same time. This encourages z_u to capture common information across domains.

Adversarial domain classifier for disentangled causal embedding: To further improve the disentanglement of domain-specific and domain-shared information in item embeddings, we propose to use a domain classifier $h_\tau(\cdot)$ to normalize the item embeddings. Specifically, we feed the items features into the domain classifier and obtain a probability score indicating the item's domain membership. Our goal is to encourage the model to learn more domain-specific features in $e_{i^x}^x$ and $e_{i^y}^y$ while learning more domain-shared features in $e_{i^x}^S$ and $e_{i^y}^S$. To achieve this, we first define the classification loss L_{cls} to encourage the domain classifier to classify items based on their domain-specific features:

$$L_{cls} = -\sum_{i^x \in \mathcal{I}^x} \log h_\tau(e_{i^x}^x) - \sum_{i^y \in \mathcal{I}^y} \log(1 - h_\tau(e_{i^y}^y))$$

where $e_{i^x}^x$ and $e_{i^y}^y$ are the domain-specific features of item i^x and i^y, respectively. To further enhance the disentanglement, we introduce an adversarial loss L_{icls}, which encourages the domain classifier to predict the domain label uniformly for domain-shared feature. L_{icls} is defined as:

$$L_{icls} = -1/[\sum_{i^x \in \mathcal{I}^x} \log h_\tau(e_{i^x}^S) + \sum_{i^y \in \mathcal{I}^y} \log(1 - h_\tau(e_{i^y}^S))]$$

By minimizing both L_{cls} and L_{icls} simultaneously, the domain-specific and domain-shared features of item embeddings can be better separated, which ultimately improves the disentanglement of the causal embeddings. Note that only L_{cls} is used to optimize $h_\tau(\cdot)$.

The total training objective is formatted by: $L = L_{re} + \lambda_1 L_{KL} + \lambda_2 L_{MI} + \lambda_3(L_{cls} + L_{icls})$, where λ_1, λ_2 and λ_3 are hyper-parameters that control the trade-off between the different terms.

3 Experiment

In this section, we conduct extensive experiments to answer the following research questions (RQ): **RQ 1:** How does CausalCDR perform in the CDR task compared to baselines? **RQ 2:** How does the causal stage affect the performance? **RQ 3:** How does different hyper-parameter settings influence the performance? **RQ 4:** Could CausalCDR learn the disentangled causal embedding?

3.1 Experimental settings

Datasets. Our experiments are conducted on four real-world datasets from Amazon [31]: Elec (Electronics), Phone, Sports, and Clothing. We follow the settings used in [3, 25] and combine the datasets into four CDR scenarios: Sports&Phone, Sports&Clothing, Elec&Clothing, and Phone&Elec.

Evaluation. We evaluate the performance of methods using the Leave-One-Out [3]. For each target user u, we generate a list of 1000 items, including 1 ground truth item and 999 negative items. We then calculate the scores of these items using proposed method, and generate a top-10 list based on the scores. And we use Hit Ratio (HR) and NDCG [42] to evaluate the performance.

Baselines. We compare CausalCDR with SOTA single-domain and cross-domain baselines. Single-Domain Methods: (1)**MF** [13] is a basic method which learns user and item embedding to reconstruct interaction matrix. (2)**NCF** [10] is a deep neural network version of MF. Cross-Domain Methods: (1)**CoNet** [12] is a translation method, which designs a cross unit to transfer knowledge between domains. (2)**DARec** [48] is a translation and disentangling mixed method, which learns common preference from all domains, and decodes interaction using different decoders. (3)**DDT** [19] is a translation method designing a latent orthogonal mapping function to transfer users preference. (4)**DML** [20] is a modified version of **DDT**, which introduces dual metric learning. (5)**ETL** [6] is a translation method, which combines the embedding from different domains to decode interaction. (6)**IPS** [21] is a alignment method, which treats domain as a type of bias and aligns different domains by de-biasd embedding. (7)**SITN** [34] is a alignment method, which aligns different domains by contrasting instance-instance and instance-cluster. (8)**CausalInt** [41] is a disentangling method, which

Table 1: Performance comparison of different methods on four CDR scenarios.

Dataset	Metrics@10	Single-Domain		Cross-Domain								
		MF	NMF	CoNet	DARec	DML	DDT	ETL	IPS	SITN	CausalInt	CausalCDR
Sports	HR	0.0352	0.0654	0.1344	0.162	0.1493	0.1514	0.0717	0.1312	0.1498	0.1493	**0.1939**
Phone	NDCG	0.0177	0.0331	0.0714	0.0912	0.0808	0.0809	0.0429	0.0689	0.0802	0.0795	**0.1118**
	HR	0.053	0.0815	0.1595	0.1895	0.171	0.1723	0.1353	0.146	0.1705	0.1688	**0.2331**
	NDCG	0.0253	0.0422	0.0857	0.0996	0.0959	0.096	0.0652	0.0752	0.0945	0.0944	**0.1314**
Sports	HR	0.0492	0.0476	0.1125	0.1267	0.1366	0.1384	0.1063	0.1197	0.138	0.1337	**0.1744**
Clothing	NDCG	0.0262	0.0223	0.0571	0.0694	0.0743	0.0749	0.0488	0.0631	0.0743	0.0711	**0.0969**
	HR	0.0477	0.0528	0.1141	0.1186	0.1517	0.1508	0.0842	0.1306	0.1509	0.1493	**0.1735**
	NDCG	0.0252	0.0261	0.0593	0.0641	0.0828	0.0824	0.0485	0.0733	0.0825	0.0821	**0.0982**
Elec	HR	0.143	0.104	0.2008	0.1913	0.2372	0.2368	0.1903	0.1007	0.2355	0.2346	**0.2562**
Clothing	NDCG	0.0818	0.0521	0.1117	0.1111	0.1393	0.1393	0.1046	0.0577	0.1381	0.1379	**0.1523**
	HR	0.0469	0.0466	0.1015	0.102	0.1396	0.1404	0.0896	0.0344	0.1399	0.1388	**0.1597**
	NDCG	0.0249	0.023	0.0528	0.0561	0.0764	0.0764	0.0475	0.0187	0.0765	0.0759	**0.0889**
Phone	HR	0.0516	0.1071	0.1161	0.2126	0.2071	0.2063	0.1383	0.1805	0.2059	0.2052	**0.2759**
Elec	NDCG	0.0254	0.0552	0.0599	0.1114	0.1095	0.1093	0.068	0.0916	0.11	0.1083	**0.1539**
	HR	0.0625	0.0992	0.177	0.1805	0.2098	0.2086	0.1504	0.1891	0.202	0.2086	**0.2398**
	NDCG	0.0326	0.0514	0.0994	0.1052	0.1214	0.1214	0.0801	0.1061	0.1157	0.1184	**0.1426**

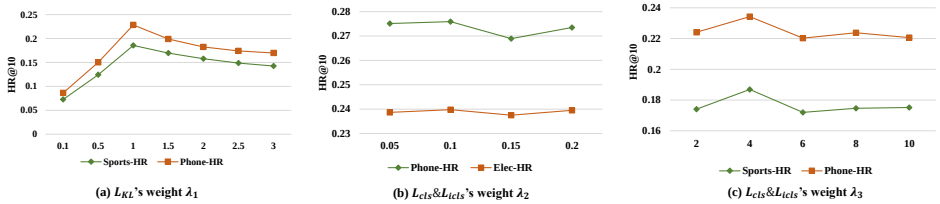


Figure 3: Performance comparison under different hyper-parameters

Table 2: Performance comparison of different variants on Elec&Clothing

Variants	Elec		Clothing	
	HR	NDCG	HR	NDCG
CausalCDR	**0.2562**	**0.1523**	**0.1597**	**0.0889**
w/o. z_u	0.2436	0.1432	0.1499	0.0822
w/o. z_u&$z_u^{x,y}$	0.1933	0.111	0.0893	0.0514
w/o. z_u&$z_u^{x,y}$&$d^{x,y}$	0.1827	0.103	0.0945	0.0542
CausalCDR-translation	0.1948	0.1114	0.1129	0.066
CausalCDR-isolated	0.1796	0.1008	0.1122	0.0651

Table 3: Metrics of K-means clustering

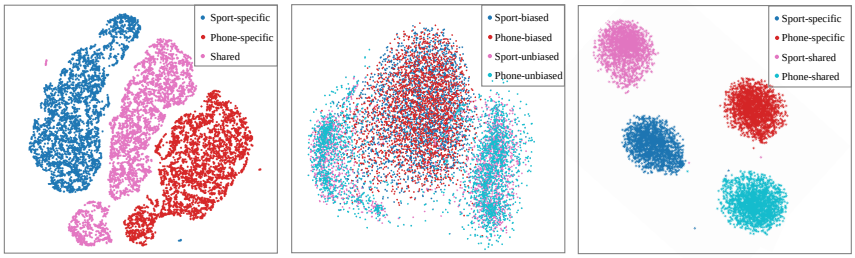
	ARI ↑			AMI ↑		
	z_u-e_u^x	z_u-e_u^y	z_u-e_u^S	z_u-e_u^x	z_u-e_u^y	z_u-e_u^S
K=4	0.3609	0.2635	0.5741	0.4799	0.3753	0.6090
K=8	0.3202	0.2856	0.4793	0.4927	0.4428	0.6103
K=16	0.3838	0.2884	0.3543	0.5714	0.4957	0.5846

disentangles the domain-shared embedding from different domains and transfers other domains' knowledge to target domain.

Implementation Details. For all methods, the common hyper-parameters are listed as follows: the initializing embedding dimension m is fixed as 128, the mini-batch size $|\mathcal{B}|$ is fixed as 512, the learning rate is fixed as 0.001, the L2 regularization coefficient is fixed as 0.0005, the cross entropy is used as reconstruction loss and the Adam optimizer [15] is used to update all parameters. We train all models with 400 epochs for convergence, and evaluate the model prediction scores every 10 epochs. And the hyper-parameters $\lambda_1, \lambda_2, \lambda_3$ of CausalCDR are set 1, 0.1, 4.

3.2 Performance Comparisons (RQ1) We conduct experiments on four CDR scenarios, and the results are shown in Table 1. From the experiment results, we get the following observations: (1) CausalCDR significantly outperforms all strong baselines. For instance, in scenario Sports&Phone, CausalCDR outperforms the best baseline DARec by 19.69% on HR and 22.59% on NDCG for the Sports domain, and outperforms it by 23.01% on HR and 31.93% on NDCG for the Phone domain. (2) In most scenarios, cross-domain methods outperform single-domain methods. However, IPS performs worse in the Elec&Clothing scenarios, possibly because it introduces noise information during alignment.

3.3 Effectiveness of Causal Stage (RQ2) To verify the effect of the causal stage, we conducted an ablation study by removing components of the causal stage and designing the following variants: (1) CausalCDR w/o. z_u removes z_u so that the causal embedding is inferred by: $e_u^x = \psi_x^C(z_u^x, d^x), e_u^y = \psi_y^C(z_u^y, d^x), e_u^S = \psi_S^C(z_u^x, z_u^y)$. (2) CausalCDR w/o. z_u&$z_u^{x,y}$ removes z_u, z_u^x, and z_u^y so that the causal embedding is inferred by: $e_u^x = \psi_x^C(X_u, d^x), e_u^y = \psi_y^C(Y_u, d^x), e_u^S = \psi_S^C(X_u, Y_u)$. (3) CausalCDR w/o. z_u&$z_u^{x,y}$&$d^{x,y}$ removes z_u, z_u^x, z_u^y, d^x, and d^y so that the causal embedding is inferred by: $e_u^x = \psi_x^C(X_u), e_u^y = \psi_y^C(Y_u), e_u^S = \psi_S^C(X_u, Y_u)$. (4) CausalCDR-translation removes z_u, z_u^x, z_u^y, d^x, d^y, and e_u^S so that the model degenerates into translation methods. The causal embedding is inferred by: $e_u^x = \psi_x^C(X_u, Y_u), e_u^y = \psi_y^C(X_u, Y_u)$. (5) CausalCDR-isolated denotes single-domain method. In this



(a) CausalCDR (b) IPS (c) CausalInt

Figure 4: Visualization of learned embeddings on Sports&Phone scenario

case, the model degenerates into isolated VAE. The causal embedding is inferred by: $e_u^x = \psi_x^C(X_u), e_u^y = \psi_y^C(Y_u)$.

The results of different variants in the four CDR scenarios are shown in Table 2. The performance decrease of the variants that remove components of the causal stage verifies the effectiveness of the proposed causal stage. Furthermore, variants that model cross-domain information outperform the single-domain variant, CausalCDR-isolated, which demonstrates that the cross-domain methods benefit from the aggregation of different domains.

3.4 Effect of hyper-parameter settings (RQ3)

In this section, we investigate the effect of hyper-parameters: λ_1 for L_{KL} weight, λ_2 for L_{MI} weight, and λ_3 for L_{cls} and L_{icls} weight. The results for Sports&Phone and Phone&Elec scenarios are shown in Figure 3. Figure 3(a) shows that the best value for λ_1 is 1. λ_1 balances the reconstruction and disentanglement of the embeddings. A too large value of λ_1 may harm the informativeness of the model, and a too small value may hurt the robustness of the probabilistic model. Figures 3(b) and (c) show that CausalCDR is insensitive to the settings of λ_2 and λ_3, as the performance fluctuations are small under different settings. The optimal values for λ_2 and λ_3 are 0.1 and 4, respectively.

3.5 Study on the learned embedding (RQ4)

Visualization of disentangled causal embedding: In this section, we validate the disentangled embedding by visualization. We use t-SNE [37] to visualize e_u^S, e_u^x, e_u^y of CausalCDR, the biased and unbiased user embedding of IPS, domain-specific and domain-shared embedding of CausalInt on Sports&Phone scenario.

For IPS shown in 4(b), both biased and unbiased embedding reside in the same embedding space, which may fail to capture domain-specific information. For CausalInt shown in Figure 4(c), the domain-specific and domain-shared embeddings exhibit clustered distributions. However, domain-shared embeddings learned form different domains are far from one another, leading to possible problems in capturing true domain-shared preferences. In contrast, the causal embeddings learned by CausalCDR, shown in Figure 4(a), are well-distributed amongst its clusters. The domain-shared embedding is inferred from both domains, avoiding the issue of domain-shared embeddings learned from different domains locating in different clusters. Furthermore, as the domain-shared embedding captures common preferences of both domains, it is close to both Sports-specific and Phone-specific.

Correlation analysis between profile embedding and causal embedding: In this section, we assess whether the profile embedding has effectively captured the user's profile. We conduct K-means clustering [18] on z_u, e_u^x, e_u^y, and e_u^S of the Sports&Phone scenario. The cluster label of z_u is treated as the true user label. The Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI) are used as metrics to evaluate the clusters of e_u^x, e_u^y, and e_u^S. ARI and AMI values range from -1 to 1, and a higher value indicates a stronger correlation between the two embeddings.

As presented in Table 3, all metrics are positive under different K settings, demonstrating that all causal embeddings are correlated with the profile embedding z_u. When two users have similar profile embeddings, they tend to exhibit similar preferences for both domain-specific and domain-shared preferences. This finding supports our claim that the isolated domain-specific preference limits the performance of CDR.

4 Related Work

4.1 Cross-domain Recommendation Cross-domain recommendation (CDR) is a solution of data sparsity by utilizing information across domains to improve recommendation performance [5, 48]. Existing CDR methods can be classified into three categories: translation, alignment, and disentanglement.

Translation methods facilitate explicit information transfer between domains [12,19]. For example, [6] proposes a dual Variational AutoEncoder model, which combines hidden embeddings from different domains to decode interactions on each domain.

Alignment methods employ isolated models and transfer information between domains via alignment mechanisms such as contrastive learning, optimal transport [21,34]. For instance, [26] trains VAE models with Gaussian mixture model priors on each domain and aligns the priors via optimal transport.

Disentanglement methods aim to separate information into domain-specific and domain-shared parts so that the shared part can be applied across all domains [4,35,41]. [3] disentangles user preferences into domain-shared and domain-specific preferences.

Our proposed method is closely related to disentangle-

ment approaches. In contrast, we incorporate causality into the CDR scenario to achieve better disentanglement. Additionally, to tackle the isolated preference issue, we model a common cause of domain-shared and domain-specific preference, the user profile information, which can be treated as an alignment component.

4.2 Disentangled representation learning Disentangled representation learning (DRL) seeks to disentangle informative factors of variation in data [27]. Recently, DRL has garnered significant attention due to favorable explainability [45, 49]. In this work, we apply a VAE-based DRL approach to recommendation systems.

VAE-based DRL methods can be broadly classified into unsupervised and supervised approaches. Unsupervised methods aim to disentangle informative factors of variation by re-weighting [11] and decomposing [2, 14] the ELBO of VAE. Despite the potential benefits, their performance is often limited due to a lack of supervision [27]. Recent research has focused more on supervised approaches [44, 46], such as group-based methods [1] and causal-based methods [17, 36]. Our proposed method builds upon CausalVAE [46], a causal-enhanced VAE-based DRL approach, but further extends this approach to model causality in CDR scenario.

DRL for Recommendation Systems DRL has recently been introduced to recommendation systems as a means of improving representation learning [24, 38], which can be divided into single-source and multi-source disentanglement methods. Single-source methods focus on disentangling the interaction information into multi-interests of users [29, 30, 40], or disentangling user conformity and true interest [50, 51]. Multi-source methods aim to disentangle different sources of information to obtain better generalization and eliminate multicollinearity [3, 39]. Different from existing studies, we exploit DRL to solve CDR problem, disentangling user's domain-shared and domain-specific preference based on user behaviors in different domains.

5 Conclusion

In this paper, we aim to address the issue of isolated domain-specific preferences in existing CDR methods and proposed CausalCDR, a novel disentangled method for cross-domain recommendation. CausalCDR models causality of variables in CDR and implements it using a generative view. Extensive experiments demonstrate the effectiveness of our proposed method in improving recommendation performance.

In future work, we plan to explore how to handle partial overlap users and non-overlap users in CDR. Additionally, we aim to develop a more general causality-based approach for CDR that can be applied to a wide range of domains beyond those studied in this paper.

Acknowledgement

This work was supported by National Natural Science Foundation of China (NSFC) under Grant Nos. 62172421 and 62072459.

References

[1] D. Bouchacourt, R. Tomioka, and S. Nowozin, *Multi-level variational autoencoder: Learning disentangled representations from grouped observations*, in AAAI'18, vol. 32, 2018.

[2] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, *Understanding disentangling in beta-vae*, arXiv, (2018).

[3] J. Cao, X. Lin, X. Cong, J. Ya, T. Liu, and B. Wang, *Disencdr: Learning disentangled representations for cross-domain recommendation*, in SIGIR, 2022.

[4] J. Cao, J. Sheng, X. Cong, T. Liu, and B. Wang, *Cross-domain recommendation to cold-start users via variational information bottleneck*, in ICDE'22, 2022.

[5] J. Chang, C. Zhang, Y. Hui, D. Leng, Y. Niu, and Y. Song, *Pepnet: Parameter and embedding personalized network for infusing with personalized prior information*, arXiv preprint, (2023).

[6] X. Chen, Y. Zhang, I. W. Tsang, Y. Pan, and J. Su, *Toward equivalent transformation of user preferences in cross domain recommendation*, TOIS, (2023).

[7] Y. Du, X. Zhu, L. Chen, Z. Fang, and Y. Gao, *Metakg: Meta-learning on knowledge graph for cold-start recommendation*, TKDE, (2022).

[8] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, *A brief review of domain adaptation*, ICDATA'20 and IKE'20, (2021).

[9] W. Gan, J. C.-W. Lin, P. Fournier-Viger, H.-C. Chao, V. S. Tseng, and S. Y. Philip, *A survey of utility-oriented pattern mining*, TKDE, (2019).

[10] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, *Neural collaborative filtering*, in WWW, 2017.

[11] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, *beta-vae: Learning basic visual concepts with a constrained variational framework*, in ICLR, 2017.

[12] G. Hu, Y. Zhang, and Q. Yang, *Conet: Collaborative cross networks for cross-domain recommendation*, in CIKM'18, 2018.

[13] M. Jamali and M. Ester, *A matrix factorization technique with trust propagation for recommendation in social networks*, in RecSys'10, 2010.

[14] H. Kim and A. Mnih, *Disentangling by factorising*, in ICML'18, 2018.

[15] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint, (2014).

[16] D. P. Kingma and M. Welling, *Auto-encoding variational bayes*, arXiv preprint, (2013).

[17] A. Komanduri, Y. Wu, W. Huang, F. Chen, and X. Wu, *Scm-vae: Learning identifiable causal representations via structural knowledge*, in Big Data, 2022.

[18] K. Krishna and M. N. Murty, *Genetic k-means algorithm*, TSMC, (1999).

[19] P. Li and A. Tuzhilin, *Ddtcdr: Deep dual transfer cross domain recommendation*, in ICDM'20, 2020.
[20] ——, *Dual metric learning for effective and efficient cross-domain recommendations*, TKDE, (2021).
[21] S. Li, L. Yao, S. Mu, W. X. Zhao, Y. Li, T. Guo, B. Ding, and J.-R. Wen, *Debiasing learning based cross-domain recommendation*, in SIGKDD'21, 2021.
[22] X. Li, X. Zhao, W. Pu, et al., *Measuring ease of use of mobile applications in e-commerce retailing from the perspective of consumer online shopping behaviour patterns*, JRCS, (2020).
[23] X. Lin, H. Chen, C. Pei, F. Sun, X. Xiao, H. Sun, Y. Zhang, W. Ou, and P. Jiang, *A pareto-efficient algorithm for multiple objective optimization in e-commerce recommendation*, in RecSys'19, 2019.
[24] Z. Lin, H. Wang, J. Mao, W. X. Zhao, C. Wang, P. Jiang, and J.-R. Wen, *Feature-aware diversified re-ranking with disentangled representations for relevant recommendation*, in SIGKDD'22, 2022.
[25] M. Liu, J. Li, G. Li, and P. Pan, *Cross domain recommendation via bi-directional transfer graph collaborative filtering networks*, in CIKM'20, 2020.
[26] W. Liu, X. Zheng, J. Su, M. Hu, Y. Tan, and C. Chen, *Exploiting variational domain-invariant user embedding for partially overlapped cross domain recommendation*, in SIGIR'22, 2022.
[27] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, *Challenging common assumptions in the unsupervised learning of disentangled representations*, in ICML'19, 2019.
[28] Y. Lu, Y. Fang, and C. Shi, *Meta-learning on heterogeneous information networks for cold-start recommendation*, in SIGKDD'20, 2020.
[29] J. Ma, C. Zhou, P. Cui, H. Yang, and W. Zhu, *Learning disentangled representations for recommendation*, NeuIPS'19, (2019).
[30] J. Ma, C. Zhou, H. Yang, P. Cui, X. Wang, and W. Zhu, *Disentangled self-supervision in sequential recommenders*, in SIGKDD'20, 2020.
[31] J. Ni, J. Li, and J. McAuley, *Justifying recommendations using distantly-labeled reviews and fine-grained aspects*, in EMNLP-IJCNLP'19, 2019.
[32] A. v. d. Oord, Y. Li, and O. Vinyals, *Representation learning with contrastive predictive coding*, arXiv preprint, (2018).
[33] I. Shenbin, A. Alekseev, E. Tutubalina, V. Malykh, and S. I. Nikolenko, *Recvae: A new variational autoencoder for top-n recommendations with implicit feedback*, in WSDM'20, 2020.
[34] G. Sun, Y. Shen, S. Zhou, X. Chen, H. Liu, C. Wu, C. Lei, X. Wei, and F. Fang, *Self-supervised interest transfer network via prototypical contrastive learning for recommendation*, arXiv preprint, (2023).
[35] H. Tang, J. Liu, M. Zhao, and X. Gong, *Progressive layered extraction: A novel multi-task learning model for personalized recommendations*, in RecSys'20, 2020.
[36] T. Q. Tran, K. Fukuchi, Y. Akimoto, and J. Sakuma, *Unsupervised causal binary concepts discovery with vae for black-box model explanation*, in AAAI'22, 2022.
[37] L. Van der Maaten and G. Hinton, *Visualizing data using t-sne.*, JMLR, (2008).
[38] W. Wang, X. Lin, L. Wang, F. Feng, Y. Ma, and T.-S. Chua, *Causal disentangled recommendation against user preference shifts*, arXiv preprint, (2023).
[39] X. Wang, H. Chen, and W. Zhu, *Multimodal disentangled representation for recommendation*, in ICME'21, IEEE, 2021.
[40] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, *Disentangled graph collaborative filtering*, in SIGIR'20, 2020.
[41] Y. Wang, H. Guo, B. Chen, W. Liu, Z. Liu, Q. Zhang, Z. He, H. Zheng, W. Yao, M. Zhang, et al., *Causalint: Causal inspired intervention for multi-scenario recommendation*, in SIGKDD'22, 2022.
[42] Y. Wang, L. Wang, Y. Li, D. He, and T.-Y. Liu, *A theoretical analysis of ndcg type ranking measures*, in Conference on learning theory, 2013.
[43] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, et al., *Mind: A large-scale dataset for news recommendation*, in ACL'20, 2020.
[44] J. Xu, Y. Ren, H. Tang, X. Pu, X. Zhu, M. Zeng, and L. He, *Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering*, in ICCV'21, 2021.
[45] M. Xu, J. Zhang, Z. Zhou, M. Xu, X. Qi, and Y. Qiao, *Learning geometry-disentangled representation for complementary understanding of 3d object point cloud*, in AAAI'21, 2021.
[46] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang, *Causalvae: Disentangled representation learning via neural structural causal models*, in CVPR'21, 2021.
[47] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, *Universal domain adaptation*, in CVPR, 2019.
[48] F. Yuan, L. Yao, and B. Benatallah, *Darec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns*, arXiv preprint, (2019).
[49] K.-Y. Zhang, T. Yao, J. Zhang, Y. Tai, S. Ding, J. Li, F. Huang, H. Song, and L. Ma, *Face anti-spoofing via disentangled representation learning*, in ECCV'20, 2020.
[50] W. Zhao, D. Tang, X. Chen, D. Lv, D. Ou, B. Li, P. Jiang, and K. Gai, *Disentangled causal embedding with contrastive learning for recommender system*, arXiv preprint, (2023).
[51] Y. Zheng, C. Gao, X. Li, X. He, Y. Li, and D. Jin, *Disentangling user interest and conformity for recommendation with causal embedding*, in WWW, 2021.
[52] F. Zhu, Y. Wang, C. Chen, J. Zhou, L. Li, and G. Liu, *Cross-domain recommendation: challenges, progress, and prospects*, arXiv preprint, (2021).