C^2DR: Robust Cross-Domain Recommendation based on Causal Disentanglement

Kong Menglin*
Central South University
School of Mathematics and Statistics
Changsha, China
212112025@csu.edu.cn

Jia Wang†
Xi'an Jiaotong-Liverpool University
School of Advanced Technology
Suzhou, China
Jia.wang02@xjtlu.edu.cn

Haiyang Zhang
Xi'an Jiaotong-Liverpool University
School of Advanced Technology
Suzhou, China
haiyang.zhang02@xjtlu.edu.cn

Yushan Pan
Xi'an Jiaotong-Liverpool University
School of Advanced Technology
Suzhou, China
yushan.pan@xjtlu.edu.cn

Muzhou Hou
Central South University
School of Mathematics and Statistics
Changsha, China
hmzw@csu.edu.cn

ABSTRACT

Cross-domain recommendation aims to leverage heterogeneous information to transfers knowledge from a data-sufficient domain (source domain) to a data-scarce domain (target domain). Existing approaches mainly ignore the modeling of users' domain specific preferences on items. We argue that incorporating domain-specific preferences from the source domain will introduce irrelevant information that fails to the target domain. Additionally, directly combining domain-shared and domain-specific information may hinder the target domain's performance. To this end, we propose C^2DR, a novel approach that disentangles domain-shared and domain-specific preferences from a causal perspective. Specifically, we formulate a causal graph to capture the critical causal relationships based on the underlying recommendation process, explicitly identifying domain-shared and domain-specific information as causal irrelevant variables. Then, we introduce disentanglement regularization terms to learn distinct representations of the causal variables that obey the independence constraints in the causal graph. Remarkably, our proposed method enables effective intervention and transfer of domain-shared information, thereby improving the robustness of the recommendation model. We evaluate the efficacy of C^2DR through extensive experiments on three real-world datasets, demonstrating significant improvements over state-of-the-art baselines. The code is available at: https://github.com/KongMLin/C2DR.

CCS CONCEPTS

• **Information systems** → **Recommender systems**;

KEYWORDS

Cross-Domain Recommendation; Knowledge Transfer ; Causal Disentanglement

*Kong Menglin and Jia Wang are co-first authors with equal contribution. This work was done when Kong Menglin was a RA at Xi'an Jiaotong-Liverpool University.
†Jia Wang is the correspondence author. This work was partly supported by the Natural Science Foundation of the Jiangsu Higher Educational Institution of China (23KJB520037) and Gusu Innovation and Entrepreneurship Leading Talents Programme (ZXL2023176).

WSDM '24, March 4–8, 2024, Merida, Mexico

ACM ISBN 979-8-4007-0371-3/24/03…$15.00
https://doi.org/10.1145/3616855.3635809

1 INTRODUCTION

Personalized recommendation systems have played a crucial role in the evolution of various online applications, including e-commerce [13], search engines [19] and conversational systems [6]. Numerous recommender models have been developed based on the use of rich information captured from historical user interactions [5]. However, the scarcity of user-item interaction records often poses a challenge in accurately identifying user preferences [11]. Additionally, traditional recommendation systems encounter difficulties for new users who have no any prior interaction history, giving rise to the prevalent cold-start problem [2]. With users increasingly participating in multiple domains, there is a potential to utilize information from different domains to alleviate the challenges of data sparsity and cold start [3]. This insight has led to the development of the Cross-Domain Recommendation (CDR) [35, 37], a rapidly growing area of research in recent years.

In the CDR family, previous works can be divided into two main branches based on their approaches to transferring knowledge between domains. (1) Transfer Learning-based approaches focus on leveraging the knowledge learned in the source domain to improve recommendation performance in the target domain [8, 26]. In particular, feature-based transfer learning [7, 39] transfers useful features by mapping or sharing the feature space between the two domains. Model-based transfer learning [1] is initially trained in the source domain and then fine-tuned or adapted to the target domain using shared information (e.g., overloaded users or items). (2) Shared

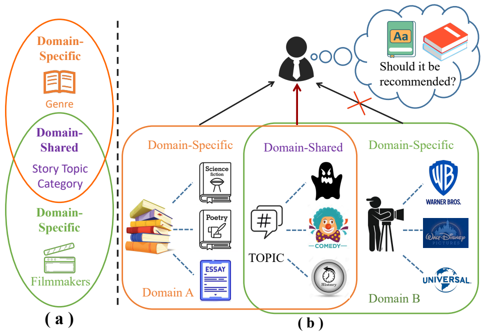

Figure 1: (a): An example of user preferences in Book and Film domain. (b): C^2DR utilizes disentangled domain-shared information and domain-specific factors for recommendations.

representation learning-based approaches aim to learn a common latent space that captures the intrinsic relationships between users and items across domains. Typically, matrix factorization-based methods [14] factorize the interaction matrices of the user-item from the source and target domains into a shared low-dimensional latent space. Some works [10, 12] utilize deep neural networks to learn complex, non-linear representations of users and items across domains. They can automatically extract and transfer high-level features, such as item content or user behavior patterns, from the source to the target domain.

In retrospect, the aforementioned technical frameworks have demonstrated promising results within their respective branches. It becomes evident that existing approaches often place excessive emphasis on the assumption of domain in-variance [1, 2], presuming that the data distribution remains unchanged in each domain. However, this assumption can be problematic, as the characteristics and distributions of user/item data can vary significantly across domains. Consequently, blindly transferring knowledge without accounting for these differences can result in transferring the irrelevant information. Motivated by recent progress [29], we further consider a challenging question in CDR: What information should be transferred? We recognize that user preferences are expressed through distinct behaviors in different domains, transferring **domain-shared information** can yield positive effects for other domains, while transferring **domain-specific information** may introduce negative transfer issues. To illustrate this point, we provide a toy example in Figure 1, consisting of two domains: "Film" and "Book," each containing several **domain-shared information** and **domain-specific information**. Intuitively, shared user preferences such as "Story Topic" and "Category" are domain-invariant. This suggests that a user's taste in these preferences is likely to be similar and stable across multiple domains. On the other hand, specific user preferences such as "Filmmakers" and "Literary Genre" provide precise intra-domain information that does not contribute to other domains. Therefore, the optimal approach is to capture and transfer the most relevant factors, specifically the **domain-shared information**, to boost recommendations in other domains.

In this paper, we explore the disentanglement of domain information from a fundamental perspective — causality, which has received little scrutiny in recommender systems. We try to elucidate how domain-share and domain-specific information serves as causal variables to influence the effects observed across and within individual domains. To achieve this, we propose an innovative causal-based framework called C^2DR for robust cross-domain recommendation. Our approach begins by constructing a causal graph that provides insights into the important causal connections that identified casual variables affect the recommendation outcomes. Furthermore, we develop multiple encoders to learn causal representations for specific items, as well as users' domain-specific interests within each domain and the domain-shared information across domains. To disentangle these causal representations, C^2DR employs a domain classifier with a gradient reverse layer (GRL) to extract domain-shared information and reweight the representation space, allowing the separation of a user's domain-specific interests within each domain. Ultimately, C^2DR ensures that the gradients of individual domain loss functions, regarding the representation of domain-shared information, are mutually orthogonal. This orthogonal property facilitates effective intervention and transfer of the causal variables, thereby enhancing the recommendation system's resilience to fluctuations in user behavior or preferences.

2 RELATED WORK

2.1 Cross-Domain Recommendation

Cross-domain recommendation (CDR) addresses data sparsity by incorporating overlapping user or item information from source domains [37]. This improves performance in the target domain even with limited data. CDR methods can be categorized into three main groups: **Collective Matrix Decomposition**: Singh et al. [25] propose the collective matrix factorization model (CMF), which factorizes rating matrices into user representations and domain-specific item representations. JCSL [18] uses a similarity matrix based on clustering and overlapping users for regularization. **Embedding and Mapping**: EMCDR [26] combines matrix factorization and Bayesian personalized ranking using a nonlinear mapping function based on a multi-layer perceptron (MLP). SSCDR [8] takes advantage of overlapping users and items from the source domain to train the mapping function. **Deep Knowledge Transfer**: CSN [**?**] combines feature maps in a high-dimensional space for bidirectional knowledge transfer. CoNet [7] introduces cross-connection units and a shared transfer matrix for fine-grained and sparse knowledge transfer. However, using a shared transfer matrix in both domains can lead to negative transfers. A recent work, MADD [33], proposes the domain disentanglement framework, which uses the attention mechanism to disentangle raw user embeddings into domain-invariant and domain-specific features. However, MADD is highly sensitive to the training strategy, which is notoriously difficult to tune. In our work, we explore the domain information disentanglement from a novel and fundamental perspective — cause-effect. We propose a causal graph for the disentanglement module design, which provides a solid theoretical foundation and enables us to develop a more robust recomender system.

2.2 Causal Inference in Recommender System

Causal inference has gained significant attention in the construction of robust machine learning models, including its application to recommendation systems. A comprehensive framework for conducting causal analysis in recommendation systems is presented in [31]. Various specific applications within recommendation systems have utilized causal inference techniques. Schnabel [21] introduces the inverse propensity score (IPS) method to handle missing-at-random data. Wang et al. [28] propose the dual robust (DR) estimator and joint optimization method to address large variance in the IPS method. Causal inference techniques have also been used to mitigate bias in recommendation systems [22, 27, 34]. Yuan et al. [32] tackle position bias in click-through rate (CTR) prediction using a multi-valued treatment approach. Sato et al. [20] discuss user self-selection bias and highlighted biased estimates even with varying sample sizes. Wei et al. [30] leverage counterfactual reasoning to eliminate popularity bias. Another relevant work [29] uses general knowledge from multiple source domains to improve performance in the target domain. Building upon these findings, we further leverage the structure causal model to develop a robust cross-domain recommendation (CDR) approach.

3 PRELIMINARY

In this research work, we consider a generalized Cross-Domain Recommendation (CDR) scenario involving two domains that have a common set of users. The data from the source domain is denoted as $\mathcal{D}^B = (\mathcal{I}^B, \mathcal{U}^B, \mathcal{Y}^B)$, and the data from the target domain is denoted as $\mathcal{D}^A = (\mathcal{I}^A, \mathcal{U}^A, \mathcal{Y}^A)$. Here, $\mathcal{I}$, $\mathcal{U}$, and $\mathcal{Y}$ represent the item set, user set, and interaction set, respectively, for each domain. Specifically, we have $\mathcal{U}^A \subseteq \mathcal{U}^B$ and $I^B \cap \mathcal{I}^A = \emptyset$. User input features are denoted as $f_u^A \in \mathcal{X}$ and $f_u^B \in \mathcal{X}$ for the target and source domains, respectively. Item input features are denoted as $f_i^A \in \mathcal{X}$ and $f_i^B \in \mathcal{X}$ in their respective domains, where $\mathcal{X}$ denotes the feature space.

Given the observed interactions matrices $\mathcal{Y}^B \in \{0,1\}^{|\mathcal{I}^B| \times |\mathcal{U}^B|}$ and $\mathcal{Y}^A \in \{0,1\}^{|\mathcal{I}^A| \times |\mathcal{U}^A|}$ in both domains, where each element A_{ij}, B_{ij} indicates whether user $i \in U$ has interacted with item $j \in I$ in the interaction set $\mathcal{Y}$, the goal of our C^2DR model is to learn disentangled representations of domain-shared information X, domain-specific information $\mathcal{X}_*$, and item representations $\mathcal{I}_*$ in both domains, where $* \in \{A, B\}$ [1]. By transferring the domain-shared representations X, we aim to enhance the recommendation performance in both domains.

4 METHODOLOGY

In this section, we first detail the causal view of the cross-domain recommendation process, followed by its rationality for disentangling domain-specified and domain-shared information. Then, we introduce the proposed C^2DR framework.

4.1 Causal Look at CDR

Causal Graph. The causal graph is a directed acyclic graph denoted as $G = V, E$, where V represents the set of variables, and E denotes the causal relations among variables [17]. In the causal graph, a capital letter (e.g., X) signifies a variable, while a lowercase letter (e.g., x) represents its observed value. An edge in the graph indicates that the ancestor node is a cause (X_A), and the successor node is an effect (Y_A). To begin, we abstract the causal graph of most existing cross-domain recommendation models, as illustrated in Figure 2. In this representation, (D) corresponds to the domain

[1]To simplify the notation, we use $*$ to indicate the operations performed in both domain A and B.

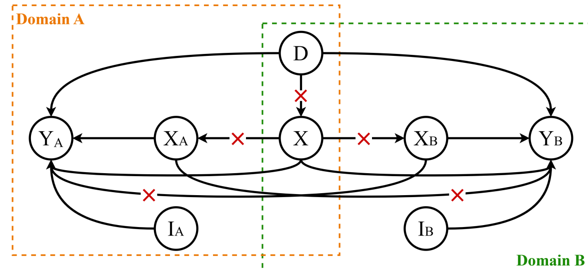

Figure 2: Cross domain recommendation casual graph. This cutting off of edges represents the elimination of causal relationships.

indicator, which is modeled as a binary variable in our task. (X) represents the embedding of the domain-shared interest (e.g., a user who enjoys love stories tends to watch romantic movies). (X_A) signifies the embedding of the user's domain-specific interest in domain A (e.g., a user who appreciates horror movies may also read a significant amount of poetry). (I_A) denotes the embedding of a specific item in domain A. Finally, (Y_A) represents the ranking score in domain A. It is important to note that the example provided here for domain A holds true for domain B as well.

Causal Relationships. Let us now elucidate the causal relationships within the causal graph, where the paths $\{D, X_A, X, I_A\} \to Y_A$ represent the direct effects of the domain indicator, user-related factors (i.e., domain-shared interest and user's domain-specific interest), and item on the ranking score. Specifically, among the variables D, X_A, X, I_A, the domain-shared information X should be independent of the domain-specific information X_A. Considering that the domain-shared information X represents high-level preferences that remain constant across domains, X should be independent of the domain indicator D. Moreover, it is crucial to avoid transferring domain-specific information to different domains (e.g., $X_B \to Y_A$), as such transfers may result in negative effects on the recommendation process and impair accurate estimation of the ranking score. To summarize the causal relationships in the causal graph, we have the following key operations:

- **Domain-shared information extraction:** We ensure that the domain-shared interest representations, denoted by X, remain independent of the domain indicator D. This independence indicates that the distribution of domain-shared interest representations does not change with different domains, i.e., $P(X|D = 0) = P(X|D = 1)$. By cutting off the causal relation $D \to X$, we achieve statistical independence between X and D (i.e., $D \perp X$).
- **Domain-specific information extraction:** We disentangle the user's domain-shared interest representations (X) from their domain-specific interests in a particular domain (X_A). Similarly, we disentangle X from the domain-specific interests in other domains (X_B). This disentanglement allows us to achieve statistical independence between X and X_A, as well as between X and X_B. By cutting off the causal relations $X \to X_A$ and $X \to X_B$, we ensure the independence of these variables (i.e., $X_A \perp X_B, X \perp X_B, X \perp X_A$).
- **Domain irrelevant information control:** To maintain the accuracy of the recommendation process, we control the transfer of domain-specific information from one domain to

another. We want to avoid negative transfer effects on the recommendation process. Therefore, we enforce the independence between the domain-specific interest in one domain (X_B) and the ranking score in another domain (Y_A). By cutting off the causal relation $X_B \longrightarrow Y_A$, we prevent unwanted interference and maintain the integrity of the recommendation process (i.e.,$X_B \perp Y_A$).

By incorporating these causal operations into the causal graph, we aim to disentangle the different factors that influence cross-domain recommendation. This disentanglement allows us to enhance the transferability of the recommendation system, improve the accuracy of ranking score estimation, and mitigate negative transfer effects. Ultimately, our approach leads to a more accurate and robust recommendation system that can adapt to various domains and provide reliable recommendations to users.

4.2 C^2DR framework

Figure 3 provides a high-level overview of C^2DR, which follows the causal graph in Figure 2 to learn the causal representation of the user's domain-specific interests within each domain and their domain-shared interests across domains. The framework consists of multiple steps. Firstly, it trains multiple encoders to acquire representations for causal variables while employing a domain classifier to extract domain-shared information. Secondly, C^2DR incorporates multiple objectives and learns a weight vector, enabling the independent treatment of causal variables ($X_A \perp X_B, X \perp X_B, X \perp X_A$). Lastly, C^2DR employs an orthogonalization constraint to mitigate negative transfer of domain-irrelevant information.

4.2.1 ***Causal representation learning.*** As introduced above, we utilize multiple encoders to learn causal embedding for various user interests across domains. Specifically, we consider the user's domain-shared interest $X \in \mathbb{R}^{|\mathcal{U}|\times d}$, which represents their interests that are common across different domains. Additionally, we examine the user's domain-specific interests in each domain, denoted as $X_A \in \mathbb{R}^{|\mathcal{U}|\times d}$ and $X_B \in \mathbb{R}^{|\mathcal{U}|\times d}$ for domains A and B, respectively. Furthermore, we incorporate item representations $I_A \in \mathbb{R}^{|\mathcal{I}^A|\times d}$ and $I_B \in \mathbb{R}^{|\mathcal{I}^B|\times d}$ for domain A and domain B, where $|\mathcal{I}^A|$ and $|\mathcal{I}^B|$ represent the number of items in each domain. To obtain these embedding, we employ the following operations:

- **User Domain-Shared Interest:** We use an encoder $X = F(emb(f_u^A \oplus f_u^B); \theta_s)$ to capture the user's domain-shared interest X. This encoder takes as input the concatenation of the domain-specific user features f_u^A and f_u^B, and applies the embedding operation $emb(\cdot)$ to obtain the embedded representation. The encoder is parameterized by θ_s.
- **User Domain-Specific Interests:** Similarly, we use the encoder $X_* = F(emb(f_u^A \oplus f_u^B); \theta_*)$ to capture the user's domain-specific interests in each domain. The encoder is parameterized by θ_*.
- **Item Representations:** The encoder $I_* = F(emb(f_i^*); \phi_*)$ is employed to capture the representation of items in each domain. This encoder takes the item features f_i^* as input and applies the embedding operation $emb_i(\cdot)$.

The function $F(\cdot)$ represents the information encoder, which can be implemented using various existing structures such as attention mechanisms [16] and graph neural networks [12]. These structures are utilized to capture users' personalized interaction information effectively. The embedding operation $emb(\cdot)$ is responsible for embedding the input features. Based on the learned representation for domain-shared information, we calculate the predicted user-item interaction result $\hat{y_s}^*$ in each domain as follows:

$$\hat{y_s}^* = Sigmoid(\mathcal{H}(X \oplus I_*; \psi_*)), \tag{1}$$

where $\oplus$ denotes the aggregation operation, $\mathcal{H}(\cdot; \psi_*)$ represents a ranking model that is implemented using a multi-layer MLP parameterized by ψ_*. Similarly, the predicted user-item interaction result $\hat{y}^*$ based on the learned representation for domain-specific information can be calculated as follows:

$$\hat{y}^* = Sigmoid(\mathcal{H}(X_* \oplus I_*; \psi_*)). \tag{2}$$

Consequently, trainable parameter set $\theta_s, \theta_*, \phi_*, \psi_*$ can be trained using supervised learning to recover the historical interactions in each domain. We use a binary cross-entropy (BCE) loss [37] to train our model:

$$\mathcal{L}_{BCE} = \sum_{*\in\{A,B\}} \sum_{\bar{y}^* \in \{\hat{y}^*, \hat{y_s}^*\}} -y^* \log \bar{y}^* - \left(1 - y^*\right) \log\left(1 - \bar{y}^*\right), \tag{3}$$

where $*$ indicates the operations performed in both domain A and B. $\bar{y}$ represents either $\hat{y_s}^*$ or $\hat{y}^*$, and y^* represents the user-item ground-truth matching score in each domain.

Note that the above training strategy has shown promise in capturing relationships within a single-domain scenario, but it faces challenges in explicitly separating domain-shared and domain-specific information in a cross-domain setting. This observation motivates us to develop a more effective training strategy for disentangling causal representations and extracting meaningful representations of user interests across different domains.

4.2.2 ***Causal representation disentanglement.*** Following the causal relationship described in subsection 4.1, it is crucial to ensure that the user's domain-shared interest representation (X) contains domain-invariant information that does not reveal specific domain information. For example, the semantic topic "Romantic" can be associated with both the "Movie" and "Film" domains. To achieve this, we draw inspiration from a domain adaptation technique [4] and label the learned representations X_A and X_B with the value '1', while labelling X with the value '0'. We then train a domain classifier $\mathcal{H}_D(\cdot; \psi_D)$ to predict the label as shown in the following equation:

$$\hat{d} = Sigmoid(\mathcal{H}_D(GRL(\cdot); \psi_D)), \tag{4}$$

where $(\cdot)$ represents the learned causal embedding of the user's domain-shared interest and domain-specific interest. Here, we utilize the gradient reversal layer (GRL) to extract the domain-shared information by confusing the domain classifier during the training process of the encoder $X = F(\cdot; \theta_s)$. The GRL acts as an identity transformation during forward propagation, i.e., $\mathrm{GRL}(X) = X$, but during backpropagation, it assigns the gradients as a negative constant, effectively reversing the sign of the gradients, i.e., $\frac{\partial \mathrm{GRL}(X)}{\partial \theta_s} = -I$. This reversal masks force the encoder $X = F(\cdot; \theta_s)$ to learn domain-invariant information, making it difficult for the

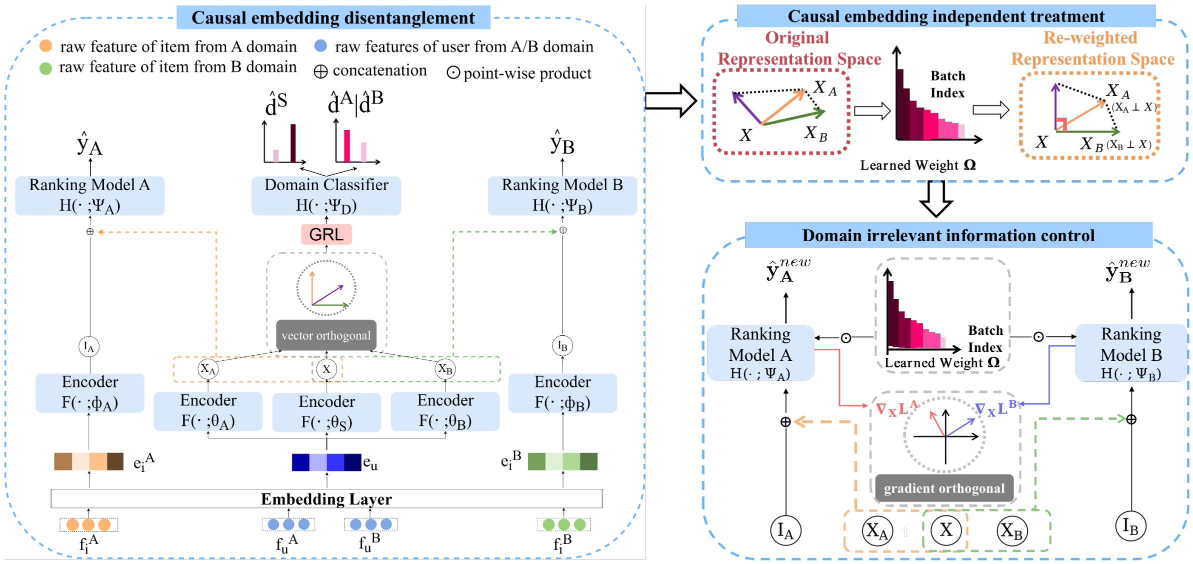

Figure 3: The training procedure of C^2DR. 1. C^2DR trains multiple encoders and employs a domain classifier with a gradient reverse layer (GRL) to extract domain-shared information effectively. 2. C^2DR learns a weight vector that re-weights the representation space to enable the independent treatment of causal variables. 3. C^2DR mitigates negative transfer of domain irrelevant information by enforcing orthogonality between the domain-shared information X and the gradients of the loss function in each domain.

domain classifier to distinguish between the source and target domains. Next, we train the domain classification loss using the following equation:

$$\mathcal{L}_{domain} = \sum_{* \in \{A,B\}} -d_* \log \hat{d}_* - 2\left(1 - d_s\right) \log \left(1 - \hat{d}_s\right), \quad (5)$$

where d_* and d_s represent the ground truth labels for the instances from domain-shared information and domain-specific information. Similarly, $\hat{d}_*$ and $\hat{d}_s$ refer to the outputs of the domain classifier when given X_A, X_B, and X as inputs, respectively.

4.2.3 *Causal embedding independent treatment*.

We treat the independence of causal embedding from both spatial geometry and statistical distribution perspective, so as to remove the undesired variable correlations, as described in subsection 4.1.

From a spatial geometry standpoint, we introduce a vector orthogonal regularizer to enforce orthogonality between the variables X, X_A, and X_B. This regularizer encourages the embedding vectors to be mutually orthogonal, promoting their independence. We calculate the vector orthogonal regularizer $\mathcal{L}_{\text{vec}}$ as follows:

$$\mathcal{L}_{\text{vec}} = \sum^{N} \text{cosine}(X, X_A) + \text{cosine}(X, X_B) + \text{cosine}(X_A, X_B), \quad (6)$$

where $\text{cosine}(\cdot, \cdot)$ represents the cosine similarity between two variables.

From a statistical distribution standpoint, since X, X_A, and X_B are continuous random variables, we achieve independence of causal embedding based on their underlying probability distribution. However, accurately estimating the probability distribution of these variables poses a challenge in our task. Therefore, we relax the strict requirement of statistical independence and instead focus on constraining the covariance between variables. For example, to encourage $X \perp X_A$, we enforce the covariance between these variables to be zero: $cov(X, X_A) = \mathbb{E}(X, X_A) - \mathbb{E}(X)\mathbb{E}(X_A) = 0$. To address this, we draw inspiration from the findings of Shen et al. [24] and introduce a learnable weight vector $\Omega \in \mathbb{R}^N$. This weight vector allows us to modify the distribution by reweighting the causal variables. By adjusting the weights, we can ensure that the covariance between the re-weighted causal variables becomes zero. Consequently, we can calculate the learnable weight Ω based on the following relationship:

$$\begin{aligned}\mathcal{L}_u(\Omega) = &\left\| \mathbf{X_A}^T \Sigma_\Omega \mathbf{X}/N - \mathbf{X_A}^T \Omega/N \cdot \mathbf{X}^T \Omega/N \right\|_F \\ &+ \left\| \mathbf{X_B}^T \Sigma_\Omega \mathbf{X}/N - \mathbf{X_B}^T \Omega/N \cdot \mathbf{X}^T \Omega/N \right\|_F \\ &+ \left\| \mathbf{X_B}^T \Sigma_\Omega \mathbf{X_A}/N - \mathbf{X_B}^T \Omega/N \cdot \mathbf{X_A}^T \Omega/N \right\|_F = 0,\end{aligned} \quad (7)$$

where N denotes the batch size, $\Sigma_\Omega \in \mathbb{R}^{N \times N}$ is a diagonal matrix with the values of the weight vector as its diagonal elements and 0 in other positions, and $|\cdot|_F$ represents the Frobenius norm. To simplify the calculation, we transform Equation (7) and optimize the learnable weight Ω by minimizing the following objective:

$$\begin{aligned}&\min_{\Omega} \mathcal{L}_u(\Omega) \\ &\text{s.t } \frac{1}{N}\sum_{i=1}^{N} \Omega_i^2 < \lambda_1, \\ &\left(\frac{1}{N}\sum_{i=1}^{N} \Omega_i - 1\right)^2 < \lambda_2, \Omega \geq \mathbf{0},\end{aligned} \quad (8)$$

where Ω_i is the value at the i-th position of Ω, and λ_1 and λ_2 are hyperparameters used to regularize the value of the weight vector. By solving this optimization problem, we can update the learnable weight vector Ω. Next, we update the ranking model $\mathcal{H}(\cdot; \psi_*)$ in the re-weighted representation space with the following objective:

$$\mathcal{L}_{new}^{*} = \sum_{i=1}^{N} \sum_{* \in \{A,B\}} \Omega_i (-y^* \log \hat{y}_{new}^{*}) - \Omega_i \left(1 - y^*\right) \log \left(1 - \hat{y}_{new}^{*}\right), \tag{9}$$

where $\hat{y}^{\text{new}}$ represents the ultimate user-item matching score based on the causal embedding of domain-shared information X and domain-specific information X_*:

$$\hat{y}_{new}^{*} = Sigmoid(\mathcal{H}(X_* \oplus X \oplus I_*; \psi_*)). \tag{10}$$

4.2.4 ***Domain irrelevant information control.*** Note that the ranking model $\mathcal{H}(\cdot; \psi_*)$ is jointly trained on both domains (Equation (9)), it is important to address the challenge of negative transfer of domain-irrelevant information. This issue becomes particularly evident when the user-item matching scores (Y_A, Y_B) of the two domains exhibit a high degree of uncorrelation. In such cases, a decrease in the loss of one domain can lead to an increase in the loss of the other domain, thereby hindering the convergence of the model. To tackle this issue, we propose the adoption of an orthogonalization constraint on the gradients of the loss function with respect to the domain-shared information (X) in both domains. The orthogonalization constraint is defined as:

$$\mathcal{L}_{grad} = \sum_{i=1}^{N} l_2^{\text{norm}} \left(\frac{\nabla_X \mathcal{L}_{new}^{A}}{\left\|\nabla_X \mathcal{L}_{new}^{A}\right\|} \cdot \frac{\nabla_X \mathcal{L}_{new}^{B}}{\left\|\nabla_X \mathcal{L}_{new}^{B}\right\|} \right), \tag{11}$$

where $l_2^{\text{norm}}(\cdot)$ denotes the L2 norm of vectors. This constraint enforces orthogonality between the domain-specific information X_B and the ranking results Y_A, as well as between X_A and Y_B, i.e., $X_B \perp Y_A$ and $X_A \perp Y_B$. By imposing orthogonality, we ensure that the gradients of the loss function with respect to the domain-shared information are orthogonal to the gradients with respect to the ranking results in both domains. Enforcing orthogonality allows us to control the flow of domain-specific and domain-shared information, enabling the model to prioritize the relevant aspects of each domain while mitigating interference from other domains. As a result, the model can focus on the relevant aspects of each domain, allowing for more robust recommendation .

4.2.5 ***Model training and evaluation.*** The totally loss function of C^2DR contains two stages. In the first stage, we update the domain-shared information encoder $F(\cdot; \theta_S)$, domain-specific information encoder $F(\cdot; \theta_*)$,item representation encoder $F(\cdot; \phi_*)$, and ranking models $\mathcal{H}(\cdot; \psi_*)$ as follows.

$$\mathcal{L}_1 = \mathcal{L}_o + \alpha \mathcal{L}_{domain} + \beta \mathcal{L}_{vec}, \tag{12}$$

where α and β are the hyperparameter. In the second stage, we learn the re-weight vector based on Equation (8), and further update the ranking models $h(\cdot; \psi_*)$ as follows.

$$\mathcal{L}_2 = \mathcal{L}_{new}^{*} + \gamma \mathcal{L}_{grad}, \tag{13}$$

where γ is the hyperparameter. The pseudo-code for the complete training process of C^2DR is shown in Appendix A.

5 EXPERIMENTS

In this section, we present the extensive experiments conducted to evaluate the performance of our proposed C^2DR model and address the following research questions:

- RQ1: Does C^2DR achieve significant performance improvements compared to existing CDR methods?
- RQ2: To what extent does each loss component in the C^2DR model contribute to overall performance improvement? Is the inclusion of each component necessary?
- RQ3: To evaluate whether the carefully designed information disentanglement regularization terms work as expected and if the learned representations satisfy the causality depicted in Figure 2.

To answer these questions, we performed the following comprehensive set of experiments and analyses.

5.1 Experimental Setup

5.1.1 Dataset. In this section, we describe the datasets used to evaluate and compare the proposed model with other models. The experiments were conducted on three large real-world datasets, each containing interaction records from two domains with a common user. We refer to the domain with fewer interaction records as the target domain and the other as the source domain. The details of the datasets used in our experiments are shown in table 1.

Table 1: Statistics of three public datasets. (avg. - average)

Dataset	**Douban**	**Huawei**	**Amazon**
# Shared Users	23,706,610	65297	87,896
#Items(Source)	54,829,040	98600	673,826
# Items(Target)	13,149,185	12615	100,164
#Instances	445,821,389	10903249	1,290,358
Avg.# clicked(Source)	94	18	39
Avg. # clicked(Target)	60	118	36

- Huawei[2]: We obtained a large-scale dataset from Huawei's 2022 cross-domain Click-Through Rate (CTR) prediction competition. It includes extensive features and comprises a 'news' source domain and an 'advertisement' target domain.
- Amazon[3]: We utilized the publicly available Amazon reviews dataset, focusing on the 'music' and 'movie' categories. Common users were extracted, treating 'movie' as the source domain and 'music' as the target domain.
- Douban[4]: This dataset was collected from Douban, encompassing three popular domains: Douban Book, Douban Movie, and Douban Music. 'Movie' was used as the source domain, and 'Book' was used as the target domain in our experiments. Please note that this dataset is not publicly accessible and has been employed in a prior study [29].

5.1.2 Baseline. For the baseline models, we include DIN [36] and its two variants as representatives of single-domain methods. Additionally, we select the following seven models as representatives of cross-domain methods:

Single-domain methods.

[2]https://www.huawei.com/
[3]http://jmcauley.ucsd.edu/data/amazon/index_2014.html
[4]https://www.douban.com/

- DIN [36]: This attention-based pooling model predicts click-through rates (CTR) effectively by capturing user interests. It serves as the baseline without knowledge-transfer modules.
- DIN (Mix): This variant of the single-domain DIN model which is trained and tested on a dataset that consists of overlapped user features from two domains.
- DIN (Finetune): This variant of the single-domain DIN model is initially pre-trained on the source domain and subsequently fine-tuned on the target domain.

Cross-domain methods : Embedding and Mapping.

- EMCDR [26]: This approach combines matrix factorization and Bayesian personalized ranking to generate latent factors and incorporates a nonlinear multi-layer perceptron (MLP) for ranking.
- PTUPCDR [39]: This method utilizes a meta mapping network as a personalized bridge function based on EMCDR.

Cross-domain methods : Deep Knowledge Transfer

- CSN [15]: This model achieves bidirectional knowledge transfer by combining feature maps in a high-dimensional space.
- CoNet [7]: This model uses cross-connection units and a shared transfer matrix for fine-grained knowledge transfer.
- MiNet [16]: This model employs a hierarchical attention mechanism to extract and transfer knowledge between domains.
- ACDR [9]: This model incorporates adversarial learning to capture both global user preferences and domain-specific user preferences across different domains.
- MADD [33]: This model utilizes the attention mechanism to construct personalized preferences by disentangling raw user behavior into domain-shared and domain-specific features.

5.1.3 Evaluation Metrics. We evaluate the performance of C^2DR and selected baseline models on different datasets with the following ranking metrics:

- AUC over the test set [38]. It is a widely used metric for CTR prediction. It reflects the probability that a model ranks a randomly chosen positive instance higher than a randomly chosen negative instance.
- RelaImpr (Relative Improvement) [23]. It calculates the percentage improvement achieved by the target model over the baseline models. It allows for a comparative analysis of the performance enhancement.

5.1.4 Hyperparameters. In our experiments, we maintain consistent settings for common hyperparameters across all methods. These settings include embedding dimension d of 64, a batch size N of 2000, a learning rate η of 0.0001, and Adam optimizer. The encoder $F(\cdot)$ is implemented as the 3-layers MLP and ranking model $\mathcal{H}(\cdot)$ is implemented as 2-layers MLP. We employ early stopping with a patience of 3 epochs for all models, and set M_1 steps and M_2 steps are 4000 and 2000, respectively. For the baseline models, we utilize the hyperparameter values reported in their original literature. Regarding C^2DR, we set α and β in Equation (12) to 0.5 and 1.5, respectively, while γ in Equation (13) is set to 1.0. These specific values were chosen based on preliminary experiments and empirical observations.

5.2 Performance Comparisons (RQ1)

We perform 10 random experiments for each model on prediction tasks across three cross-domain datasets. The reported results represent the mean performance averaged over these experiments. To ensure fairness and validity, we adjust the structure of the baseline models to match the complexity of the model C^2DR. Based on the experimental results presented in Table 2, we make the following observations.

The semantic correlation of data in different domains has a significant impact on the performance of CDR models. Traditional methods demonstrate significant improvement in datasets with strong inter-domain correlation. Specifically, EMCDR and PTUPC models outperform the baseline model by 0.07% and 1.18%, respectively. However, in the Amazon dataset characterized by weak domain correlation, traditional methods encounter difficulties in establishing meaningful connections between the two domains, resulting in performance degradation 1.08% for EMCDR and 1.18% for PTUPCDR. In contrast, The proposed C^2DR shows better performance than traditional baselines in both dataset, which validates the importance of considering domain differences and transferring domain-specific knowledge to achieve robust recommendations.

The improvement of model performance in the source domain is not as significant as in the target domain. For example, the RelImp of CoNet on Huawei_Ad is +0.21%, but it remains slightly decreases -0.52% on Huawei_News. This observation can be attributed to the fact that the source domain contains more interaction records about the domain-shared representation of user interest. The model pretrained on the source domain provides a better initial value for training the target domain. MADD and C^2DR achieve relatively consistent performance improvements in both the source and target domains in the three data sets. This is due to the design of a representation decoupling module that facilitates the extraction of complete domain-shared representations of user interest from both domains.

Domain information disentanglement plays a crucial role in achieving robust CDR performance. It is evident that the performance of CSN, CoNet, and MiNet varies across different datasets. For example, CSN exhibits 0.10% increase in Huawei_News and 0.14% increase in Huawei_Ad. Similarly, CoNet shows a 0.52% decrease on Huawei_News but a 0.21% increase on Huawei_Ad. Also, their performance notably declines in datasets with poor domain correlations. In Amazon_Music, CSN and CoNet experience performance decreases of 1.30% and 1.40%, respectively, while they experience decreases of 0.18% and 0.27% in Amazon_Movies, respectively. This discrepancy can be attributed to CSN and CoNet employing the same coefficient/matrix structure for domain information transfer. On the other hand, ACDR, MADD and our proposed C^2DR consistently enhance performance across all datasets, as they model the domain-shared and domain-specific representations through representation disentanglement. In particular, C^2DR achieves the best overall performance, highlighting the advantages of causal modeling.

5.3 Discussion of Model Variants (RQ2)

To evaluate the efficacy of each loss component in C^2DR model, we perform ablation experiments on the Huawei and Amazon datasets.

Table 2: Performance comparison of different methods for CDR scenarios on three datasets. The best-performing method is highlighted in bold font. An asterisk (∗) indicates a p-value < 0.05 for a one-tailed t-test, indicating statistically significant differences. The underlining denotes sub-optimal performance. The reported results are averaged over 10 independent repetitions across all datasets.

Datasets		Metircs	Single-domain Methods			Cross-domain Methods							Ours
			DIN	DIN (Mix)	DIN (Finetune)	EMCDR	PTUPCDR	CSN	CoNet	MiNet	ACDR	MADD	C^2DR
Huawei	Advertisement	AUC	0.7793	0.7847	0.7823	0.7695	0.7795	0.7826	0.7797	0.7799	0.7881	0.7871	**0.7892***
		RelImp		1.93%	1.07%	0.07%	1.18%	0.14%	0.21%	2.01%	3.11%	2.79%	**3.54%**
	News	AUC	0.7859	0.7836	0.7857			0.7862	0.7844	0.7905	0.7934	0.7939	**0.7942***
		RelImp		-0.80%	-0.07%			0.10%	-0.52%	1.61%	2.62%	2.80%	**2.90%**
Amazon	Music	AUC	0.6852	0.6823	0.6865	0.6832	0.6835	0.6828	0.6826	0.6875	0.6889	0.6885	**0.6898***
		RelImp		-1.57%	0.70%	-1.08%	-0.92%	-1.30%	-1.40%	1.24%	2.00%	1.78%	**2.48%**
	Movie	AUC	0.7214	0.7196	0.7219			0.7210	0.7208	0.7238	0.7249	0.7254	**0.7258**
		RelImp		-0.81%	0.23%			-0.18%	-0.27%	1.08%	1.58%	1.81%	**1.99%**
Douban	Book	AUC	0.7562	0.7578	0.7582	0.7576	0.7588	0.7554	0.7542	0.7585	0.759	0.7582	**0.7611***
		RelImp		0.62%	0.78%	0.55%	1.01%	-0.31%	-0.78%	0.90%	1.09%	0.78%	**1.91%**
	Movie	AUC	0.7802	0.7732	0.7856			0.7866	0.7906	0.7894	0.7894	0.7915	**0.792**
		RelImp		-2.50%	1.93%			2.28%	3.71%	3.28%	3.28%	4.03%	**4.21%**

The DIN (single-domain) model serves as the baseline for comparison. We compared the five special cases: C^2DR (1) without $\mathcal{L}_u$ that ensures the independence of domain shared and domain-specific information; (2) without $\mathcal{L}_{vec}$ that enforces orthogonality between domain-shared and domain-specific information in vector spaces; (3) without $\mathcal{L}_{domain}$ which involves the constraint of the domain-shared/specific information classifier; and (4) without $\mathcal{L}_{grad}$ that blocks the transfer of domain irrelevant information.

In Figure 4, we observe that all loss components contribute to the recommendation performance. Notably, removing $\mathcal{L}u$ and $\mathcal{L}vec$ has a more detrimental effect on the model's performance compared to removing $\mathcal{L}domain$ and $\mathcal{L}grad$. Similarly, removing $\mathcal{L}domain$ has a more significant negative impact than removing $\mathcal{L}grad$. These findings confirm the benefits of explicit representation disentanglement in improving CDR performance. Additionally, removing $\mathcal{L}u$ results in significantly poorer performance compared to removing $\mathcal{L}vec$. This result further confirms the importance of effectively constraining the dependence between domain-shared and domain-specific information.

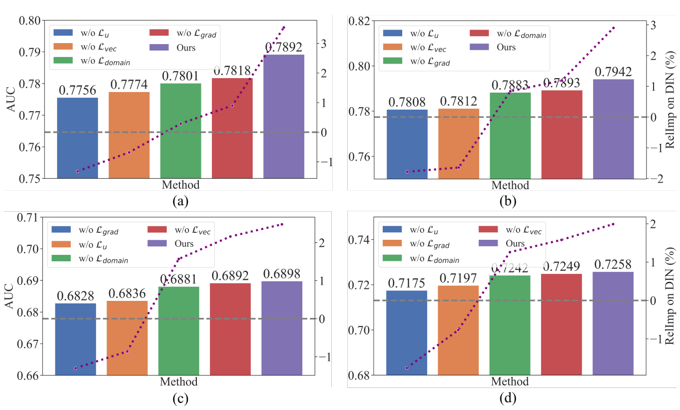

Figure 4: The impact of each loss component of C^2DR on the final performance. (a): Results on the Huawei_News (target domain). (b): Results on the Huawei_Ad (source domain) . (c): Results on the Amazon_Movie (target domain). (d): Results on the Amazon_Music (source domain). w/o '*' represents the results on the test set for the variant with the '*' loss component removed from C^2DR. The purple dotted dash line in the figure represents the Relative Improvement (RelImpr) of the model performance compared to DIN (depicted by the gray horizontal dashed line).

5.4 Effectiveness of Disentanglement (RQ3)

In this subsection, we compare the effectiveness of C^2DR's causal-based domain information disentanglement approach with CoNet in learning domain-specific representations for users. Specifically, we evaluate the performance of cross-domain recommendation in each domain using the extracted representations X_A and X_B, which represent the user's domain-specific interests in the target and source domains, respectively.

The analysis is performed on the Huawei and Douban dataset. The results presented in Table 3 and 4 reveal the interesting findings. One can see that the domain-specific representations learned by CoNet exhibit similar recommendation capacity in both domains (i.e., using X_A in the target domain and X_B in the source domain). However, they failed to achieve better results in their respective domains. In contrast, the domain-specific representations X_A and X_B learned by C^2DR demonstrate strong performance within their respective domains and exhibit weak performance when applied to the opposite domains (i.e., using X_A in the source domain and X_B in the target domain). These findings highlight the capability of C^2DR's causal-based domain information disentanglement approach in learning domain-specific representations, enabling precise and domain-aware recommendations.

6 CONCLUSION

In this paper, we have proposed a novel causal-based framework, C^2DR, to address the challenges of Cross-Domain Recommendation

Table 3: AUC performance of CDR based on domain-specific information on Huawei dataset.

Input	C^2DR		CoNet	
	Target domain	Source domain	Target domain	Source domain
X_A	**0.7892**	0.71	0.7799	0.72
X_B	0.71	**0.7942**	0.73	0.7874

Table 4: AUC performance of CDR based on domain-specific information on the Douban dataset.

Input	C^2DR		CoNet	
	Target domain	Source domain	Target domain	Source domain
X_A	**0.761**	0.72	0.754	0.73
X_B	0.73	**0.792**	0.74	0.791

(CDR) by disentangling domain information from a causal perspective. Our approach leverages causal relationships to identify and transfer the most relevant factors, specifically the domain-shared information, for improving recommendations in different domains. The proposed C^2DR model constructs a causal graph to capture important causal connections and utilizes multiple encoders to learn causal representations for domain-specific interests and domain-shared information. By employing a domain classifier and sample re-weighting techniques, C^2DR disentangles domain-specific and domain-shared information, enabling selective transfer and enhancing recommendation performance. Experimental results on real-world datasets demonstrate the effectiveness of C^2DR compared to state-of-the-art methods, while offering insights into how causal variables influence recommendation outcomes. Our research contributes a novel perspective to the field of cross-domain recommendation and opens avenues for further exploration of causal graph formulation and disentangled representations for multi-domain recommendation systems.

REFERENCES

[1] Hamed Bonab, Mohammad Aliannejadi, Ali Vardasbi, Evangelos Kanoulas, and James Allan. 2021. Cross-market product recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 110–119.

[2] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-Domain Recommendation to Cold-Start Users via Variational Information Bottleneck. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. 2209–2223. https://doi.org/10.1109/ICDE53745.2022.00211

[3] Wenjing Fu, Zhaohui Peng, Senzhang Wang, Yang Xu, and Jin Li. 2019. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[4] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The journal of machine learning research* 17, 1 (2016), 2096–2030.

[5] Xiaotian Han, Chuan Shi, Senzhang Wang, S Yu Philip, and Li Song. 2018. Aspect-Level Deep Collaborative Filtering via Heterogeneous Information Networks.. In *IJCAI*.

[6] Harald, Steck, Xiwang, Yang, Yong, Liu, Yang, and Guo. 2014. A survey of collaborative filtering based social recommender systems. *Computer communications* (2014).

[7] G. Hu, Z. Yu, Y. Qiang, and H. U. Guang-Neng. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. *ACM* (2018).

[8] S. K. Kang, J. Hwang, D. Lee, and H. Yu. 2019. Semi-Supervised Learning for Cross-Domain Recommendation to Cold-Start Users. In *the 28th ACM International Conference*.

[9] Pan Li, Brian Brost, and Alexander Tuzhilin. 2022. Adversarial Learning for Cross Domain Recommendations. *ACM Trans. Intell. Syst. Technol.* 14, 1, Article 5 (nov 2022), 25 pages. https://doi.org/10.1145/3548776

[10] P. Li and A. Tuzhilin. 2020. DDTCDR: Deep Dual Transfer Cross Domain Recommendation. In *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining*.

[11] Xixun Lin, Jia Wu, Chuan Zhou, Shirui Pan, Yanan Cao, and Bin Wang. 2021. Task-Adaptive Neural Process for User Cold-Start Recommendation. In *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, NY, USA, 1306–1316. https://doi.org/10.1145/3442381.3449908

[12] M. Liu, G. Li, and P. Pan. 2020. Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*.

[13] J. Mai, Y. Fan, and Y. Shen. 2009. A Neural Networks-Based Clustering Collaborative Filtering Algorithm in E-Commerce Recommendation System. In *International Conference on Web Information Systems Mining*.

[14] J. Meng, C. Peng, N. J. Yuan, X. Xing, and S. Yang. 2016. Little Is Much: Bridging Cross-Platform Behaviors through Overlapped Crowds. *AAAI Press* (2016).

[15] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert. 2016. Cross-stitch Networks for Multi-task Learning. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[16] Wentao Ouyang, Xiuwu Zhang, Lei Zhao, Jinmei Luo, Yu Zhang, Heng Zou, Zhaojie Liu, and Yanlong Du. 2020. Minet: Mixed interest network for cross-domain click-through rate prediction. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 2669–2676.

[17] J. Pearl. 2000. Causality: Models, reasoning, and inference, second edition. *Cambridge University Press* (2000).

[18] Dimitris Rafailidis and Fabio Crestani. 2016. Top-N Recommendation via Joint Cross-Domain User Clustering and Similarity Learning. *Springer, Cham* (2016).

[19] Richardson, Matthew, Dominowska, Ewa, and Robert. 2007. Predicting clicks: Estimating the click-through rate for new ads. In *the 16th international conference*.

[20] Y. Saito. 2020. Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. In *SIGIR '20: The 43rd International ACM SIGIR conference on research and development in Information Retrieval*.

[21] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. (2016).

[22] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.

[23] Q. Shen, W. Tao, J. Zhang, H. Wen, Z. Chen, and Q. Lu. 2021. SAR-Net: A Scenario-Aware Ranking Network for Personalized Fair Recommendation in Hundreds of Travel Scenarios. (2021).

[24] Zheyan Shen, Peng Cui, Jiashuo Liu, Tong Zhang, Bo Li, and Zhitang Chen. 2020. Stable Learning via Differentiated Variable Decorrelation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*.

[25] A. P Singh and G. J Gordon. 2008. Relational learning via collective matrix factorization. In *Acm Sigkdd International Conference on Knowledge Discovery Data Mining*.

[26] M. Tong, H. Shen, X. Jin, and X. Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*.

[27] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2020. Click" is not equal to "like": Counterfactual recommendation for mitigating clickbait issue. *arXiv preprint arXiv:2009.09945* (2020).

[28] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly robust joint learning for recommendation on data missing not at random. In *International Conference on Machine Learning*. PMLR, 6638–6647.

[29] Yichao Wang, Huifeng Guo, Bo Chen, Weiwen Liu, Zhirong Liu, Qi Zhang, Zhicheng He, Hongkun Zheng, Weiwei Yao, Muyu Zhang, Zhenhua Dong, and Ruiming Tang. 2022. CausalInt: Causal Inspired Intervention for Multi-Scenario Recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data*.

[30] T. Wei, F. Feng, J. Chen, Z. Wu, J. Yi, and X. He. 2021. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. (2021).

[31] Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, and Xiao-Hua Zhou. 2022. On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges. In *Proceedings of the International Joint Conference on Artificial Intelligence, Vienna, Austria*. 23–29.

[32] B. Yuan, Y. Liu, J. Y. Hsia, Z. Dong, and C. J. Lin. 2020. Unbiased Ad Click Prediction for Position-aware Advertising Systems. In *14th ACM Conference on Recommender Systems*.

[33] Xinyue Zhang, Jingjing Li, Hongzu Su, Lei Zhu, and Heng Tao Shen. 2023. Multi-Level Attention-Based Domain Disentanglement for BCDR. *ACM Trans. Inf. Syst.* 41, 4 (mar 2023), 24. https://doi.org/10.1145/3576925

[34] Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal intervention for leveraging popularity bias in recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 11–20.

[35] Yi Zhao, Chaozhuo Li, Jiquan Peng, Xiaohan Fang, Feiran Huang, Senzhang Wang, Xing Xie, and Jibing Gong. 2023. Beyond the overlapping users: cross-domain recommendation via adaptive anchor link learning. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

[36] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1059–1068.

[37] F. Zhu, Y. Wang, C. Chen, J. Zhou, L. Li, and G. Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. (2021).

[38] J. Zhu, J. Liu, S. Yang, Q. Zhang, and X. He. 2021. Open Benchmarking for Click-Through Rate Prediction.. In *Conference on Information and Knowledge Management*.

[39] Y. Zhu, Z. Tang, Y. Liu, F. Zhuang, R. Xie, X. Zhang, L. Lin, and Q. He. 2021. Personalized Transfer of User Preferences for Cross-domain Recommendation. (2021).