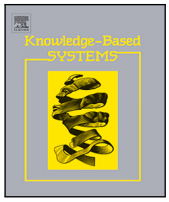
Knowledge-Based Systems 304 (2024) 112529

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



A privacy-preserving framework with multi-modal data for cross-domain recommendation

Li Wang [a], Lei Sang [b], Quangui Zhang [c], Qiang Wu [a], Min Xu [a,*]

[a] School of Electrical and Data Engineering, University of Technology Sydney, 15, Broadway, Ultimo, Sydney, 2000, NSW, Australia
[b] School of Computer Science and Technology, Anhui University, 111, Jiulong Road, Economic and Technological Development District, Hefei, 230601, China
[c] School of Artificial Intelligence, Chongqing University of Arts and Sciences, 319, Honghe Avenue, Yongchuan District, Chongqing, 402160, China

ARTICLE INFO

Keywords:
Privacy-preserving
Multi-modal
Disentanglement
Contrastive learning
Cross-domain recommender systems

ABSTRACT

Cross-domain recommendation (CDR) aims to enhance the recommendation accuracy in a target domain with sparse data by leveraging rich information in a source domain, thereby addressing the data-sparsity problem. Some existing CDR methods highlight the advantages of extracting domain-common and domain-specific features to learn comprehensive user and item representations. However, these methods cannot effectively disentangle these components, as they often rely on simple user-item historical interaction information (such as ratings, clicks, and browsing), neglecting the rich multi-modal features. In addition, they do not protect user-sensitive data from potential leakage during knowledge transfer between domains. To address these challenges, we propose a **P**rivacy-**P**reserving Framework with **M**ulti-**M**odal Data for **C**ross-**D**omain **R**ecommendation, called P2M2-CDR. Specifically, we first design a multi-modal disentangled encoder that utilizes multi-modal information to disentangle more informative domain-common and domain-specific embeddings. Furthermore, we introduce a privacy-preserving decoder to mitigate user privacy leakage during knowledge transfer. Local differential privacy (LDP) is used to obfuscate disentangled embeddings before the inter-domain exchange, thereby enhancing privacy protection. To ensure both consistency and differentiation among these obfuscated disentangled embeddings, we incorporate contrastive learning-based domain-inter and domain-intra losses. Extensive experiments conducted on six CDR tasks from two real-world datasets demonstrate that P2M2-CDR outperforms other state-of-the-art single- and cross-domain baselines. The code is available at https://github.com/Lili1013/P2M2-CDR.

1. Introduction

Cross-domain recommendation (CDR) has emerged as a crucial strategy for addressing the persistent issue of data sparsity in recommendation systems by transferring informative knowledge across related domains [1,2]. The inherent data-sparsity problem arises when the target domain lacks sufficient user-item interaction information, hindering the ability to provide accurate and personalized recommendations. To address this challenge, various methods [2–10] have been proposed to enhance the performance of CDR. These methods leverage the rich knowledge in the source domain to complement sparse information in the target domain. Collective matrix factorization (CMF)-based methods [3,4] concentrate on creating shared user and item representations in different domains by matrix factorization technology. In contrast, embedding-based methods [5,6] train separate encoders to learn embeddings and subsequently employ mapping techniques to project user and item embeddings onto a shared space. Dual knowledge transfer-based methods [2,7–9] are dedicated to bidirectional knowledge transfer, while graph neural networks (GNNs)-based methods [10] leverage high-order collaborative information to enhance performance. However, these approaches often assume that the users in different domains have similar interests. They tend to learn user representations by relying on shared knowledge, thereby overlooking the possibility that users have diverse preferences across different domains. This means that they focus solely on domain-common information while ignoring domain-specific information.

It is crucial to separate domain-common and domain-specific knowledge to ensure that they contain additional semantic information. In recent years, several methods [9,11–13] have emerged that simultaneously capture both domain-common and domain-specific knowledge

* Corresponding author.
E-mail addresses: li.wang-13@student.uts.edu.au (L. Wang), sanglei@ahu.edu.cn (L. Sang), zhqgui@cqwu.edu.cn (Q. Zhang), Qiang.Wu@uts.edu.au (Q. Wu), Min.Xu@uts.edu.au (M. Xu).

https://doi.org/10.1016/j.knosys.2024.112529
Received 27 December 2023; Received in revised form 22 August 2024; Accepted 13 September 2024
Available online 19 September 2024

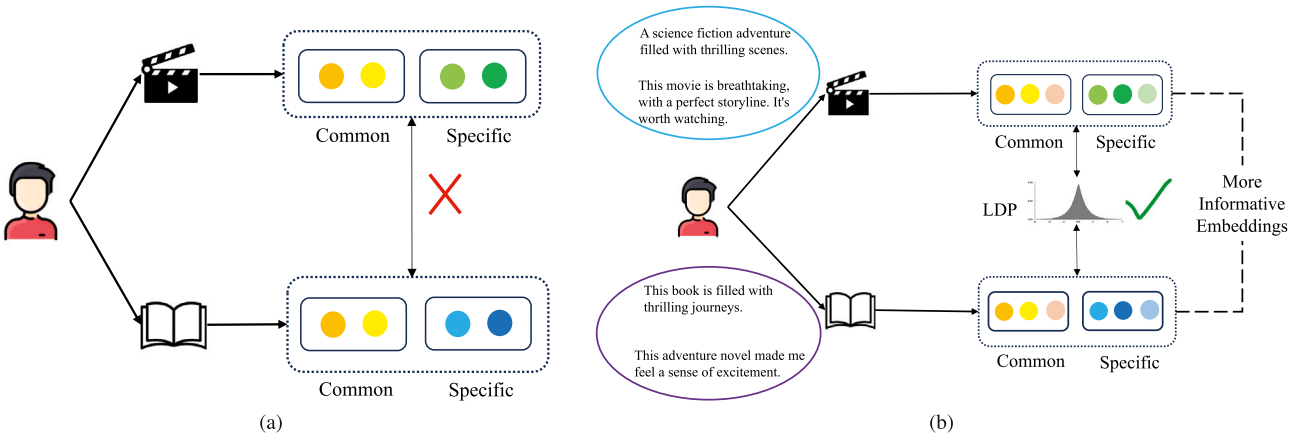


Fig. 1. The illustration of earlier CDR methods (a) and P2M2-CDR (b). In comparison to earlier CDR methods, P2M2-CDR considers (1) introducing multi-modal data (user-item interaction history, review texts, and textual features) to disentangle more informative domain-common and domain-specific features and (2) utilizing local differential privacy technology to protect user privacy.

across different domains. As depicted in Fig. 1(a), these methods typically employ user-item interaction histories to design methods that disentangle domain-common and domain-specific features. Subsequently, these features are aggregated to facilitate the learning of user and item representations. For example, DisenCDR [11] introduces two mutual information-based regularizers to effectively disentangle domain-common and domain-specific components. Similarly, DCCDR [12] utilizes contrastive learning (CL) to decouple these two components. Additionally, DIDA-CDR [14] employs domain classifiers to disentangle domain-common, domain-specific, and domain-independent features, aiming to learn more comprehensive user representations. Although these methods have demonstrated improved performance, they have two limitations.

(1) **Limitation in Domain Disentanglement.** They depend solely on user-item historical interaction information (such as ratings, clicks, and browsing) to extract domain-common and domain-specific features. However, the recommendation accuracy may decline because only utilizing these simple user-item interaction data does not adequately decouple these features.

(2) **Limitation in Privacy Protection.** These CDR models assume that the disentangled embeddings between different domains are public and shared. However, they do not address the user privacy problem during knowledge transfer between domains. Because these disentangled embeddings are linked to sensitive user data, such as user-item ratings, sharing them increases the risk that external participants and potential attackers could infer private user information. Moreover, in real-world scenarios, these domains might be operated by different organizations. Sharing these embeddings could potentially violate business privacy policies.

Based on these limitations, in this study, we primarily address two challenges in designing a privacy-preserving CDR system. **CH1.** How can we effectively decouple the more informative domain-common and domain-specific embeddings to learn comprehensive user and item representations? **CH2.** How can we prevent the leakage of user-related information when learning and transferring disentangled embeddings? To address these challenges, we introduce P2M2-CDR, a privacy-preserving framework for CDR, as illustrated in Fig. 1(b). This framework leverages multi-modal data to disentangle domain-common and domain-specific features efficiently. In addition, we incorporate local differential privacy (LDP) [15] to mitigate the risk of user privacy leakage during knowledge transfer.

To address the first challenge, we propose a multi-modal disentangled encoder that incorporates multi-modal information to disentangle more informative domain-common and domain-specific embeddings effectively. Multi-modal information provides a more comprehensive description of users and items [15,16]. By integrating these diverse sources, the model provides a deeper understanding of user' behaviors and preferences. Consequently, it has become more proficient in distinguishing between common and specific characteristics of users and items across different domains. Specifically, for both the source and target domains, we initially learn item ID embeddings from LightGCN [17], which is a model that captures high-order collaborative relationships in the user-item interaction graph. We then combine these item ID embeddings with the item textual features obtained from the pre-trained sentence-transformer model [18] to obtain item representations. Simultaneously, we incorporate user ID embeddings learned from LightGCN and user review text embeddings acquired from sentence-transformer to effectively decouple domain-common and domain-specific embeddings. Moreover, we employ self-supervised techniques, such as the feature dropout strategy, to generate augmented disentangled embeddings.

To address the second challenge, we propose a privacy-preserving decoder that leverages LDP techniques to safeguard user-related data during the transfer of disentangled embeddings. Initially, we employ LDP to obscure domain-common and domain-specific embeddings before inter-domain sharing, thereby safeguarding user privacy from participants and external threats. Subsequently, to ensure the alignment and separation of these obfuscated decoupled embeddings, we design CL-based domain-inter and domain-intra losses. Finally, we aggregate the obfuscated domain-common and domain-specific embeddings to learn comprehensive user preferences.

In summary, the proposed model makes the following contributions:

- We propose a privacy-preserving framework for CDR, that effectively utilizes multi-modal data to generate more informative disentangled embeddings, and protects user privacy when transferring knowledge across domains. This approach solves key challenges that traditional CDR methods face.
- Unlike existing disentanglement-based CDR methods that overlook multi-modal data, we propose a multi-modal disentangled encoder. This encoder utilizes multi-modal information, including user-item interaction data, user review texts, and item textual features, to disentangle more informative domain-common and domain-specific embeddings (for **CH1**).
- We propose a privacy-preserving decoder that employs LDP techniques to prevent the leakage of user privacy information during the transfer of disentangled embeddings (for **CH2**). This effectively addresses the privacy concerns prevalent in existing CDR approaches.
- We conducted extensive experiments on six CDR tasks from two large-scale real-world datasets, Amazon and Douban. Comprehensive results demonstrate the effectiveness of P2M2-CDR compared with some state-of-the-art baselines.

2. Related work

In this section, we briefly review some representative studies related to our work, which are classified into three categories: Cross-Domain Recommendation, Disentangled Representation Learning in Recommendation, and Privacy-Preserving Recommendation. Each category includes important studies in their respective fields, providing a foundation and context for our research.

2.1. Cross-domain recommendation

CDR aims to address data-sparsity and cold-start problems by transferring rich information across diverse domains or platforms, thereby offering more accurate and personalized recommendation services. Mainstream CDR methods are generally categorized into content-based and embedding-based transfer methods [1]. Content-based CDR utilizes content features, such as user profiles [19], review texts [6], text information [20], and tags [21], to connect multiple domains by computing the similarities between users or items across domains. Tan et al. [20] propose a Bayesian hierarchical model that transfers user preferences, modeled using Latent Dirichlet Allocation (LDA) with documents, across domains. CATN [6] aims to solve the cold-start problem of the CDR by transferring aspect-level user preferences learned from review texts. Embedding-based CDR methods first leverage representation learning techniques, such as matrix factorization [22], neural collaborative filtering [23], and graph representation learning [17], to learn the latent embeddings of users or items. These embeddings are then transferred based on similar users or items. DTCDR [8] employs NeuMF or DMF to obtain rating embeddings and uses the Doc2vec model to learn document embeddings. Subsequently, it uses multi-task learning (MTL) to transfer user preferences based on common users across domains. GA-DTCDR [10] utilizes GNNs to learn user embeddings in different domains and proposes an element-wise attention network to fuse the embeddings of overlapping users.

The proposed method belongs to the second category. First, we learn the disentangled embeddings in each domain and then transfer the domain-common embeddings across domains. Although previous methods have achieved good performance, they typically assume that data across all domains are openly shared, potentially overlooking user privacy concerns. In this study, the primary focus is to mitigate the risk of user privacy leakage by utilizing LDP techniques.

2.2. Disentangled representation learning in recommendation

Disentangled representation learning in recommendation aims to decompose user characteristics or item features into distinct and independent parts, providing an effective way to enhance the robustness and interpretability of models. Disentangled representation learning has been applied to generative recommendations [24], causal recommendations [25], and graph recommendations [26]. The authors of [24] propose a model called MacridVAE, which is a disentangled variational auto-encoder that achieves both macro disentanglement of high-level concepts and micro disentanglement of isolated low-level factors. DICE [25] constructs cause-specific data based on causal effects and disentangles user and item embeddings into interest and conformity components. DGCF [26] learns disentangled representations that capture fine-grained user intent from a user-item interaction graph.

Recently, disentangled representation learning is applied to CDR. Disentanglement-based CDR methods first learn domain-common and domain-specific embeddings and then transfer domain-common embeddings across domains. According to the disentangling techniques, these methods can be categorized as VAE-based [11] or other approaches. VAE-based models employ reconstruction loss within the Evidence Lower Bound (ELBO) along with additional regularizers, which serve as disentanglement loss, to effectively learn disentangled embeddings. For example, DisenCDR [11] disentangles user preferences into domain-specific and domain-shared information by using two mutual information-based regularizers and transfers domain-shared information across both domains. Other approaches tend to use CL [12,27], adversarial learning [14], and causal learning [28] to learn disentangled representations. DCCDR [12] proposes mutual information-based CL objectives to disentangle domain-invariant and domain-specific embeddings. C^2DR [28] disentangles domain-shared and domain-specific embeddings based on causal graphs.

Our method aims to utilize CL to disentangle domain-common and domain-specific embeddings. Although the above approaches have demonstrated effectiveness, they neglect multi-modal information when disentangling domain-common and domain-specific features, resulting in suboptimal recommendation performance. Our approach effectively leverages multi-modal features to overcome this limitation.

2.3. Privacy-preserving recommendation

Privacy-preserving recommendation systems [29–32] are designed to offer personalized recommendations while protecting user privacy. The existing privacy-preserving recommendation methods can be classified into two categories: Federated Learning (FL)-based methods and LDP-based approaches. The former focuses on training data locally on each user device without sharing private data. For example, FedMF [33] extends matrix factorization techniques to a federated setting, thereby allowing private and efficient collaborative filtering. By contrast, LDP-based models, aim to protect user privacy by adding noise to the original rating data or representations. For instance, authors in [34] propose a novel approach to enhance privacy in recommendation systems by integrating matrix factorization with LDP techniques. Recently, these technologies have gained popularity in CDR [35–38] to ensure the protection of users' sensitive information during knowledge transfer. For instance, FedCDR [36] learns user and item embeddings through individual models trained on personal devices and then uploads the weights to the central server to protect the user's sensitive data. Other approaches, such as PriCDR [37], CCMF [35], and PPGenCDR [39], employ LDP techniques to construct a protected rating matrix in the source domain and subsequently transfer the perturbed rating matrix to the target domain.

Our study also leverages the LDP technology to protect user privacy when transferring knowledge across domains. Although FL-based methods can achieve good performance, they often face high computation and communication demands [40,41]. In addition, the aforementioned LDP-based approaches are single-target, focusing on transferring knowledge from a source domain with rich data to improve the recommendation quality in a target domain with sparse data. In some scenarios, both the source and target domains contain relatively rich information, such as ratings, review texts, user characteristics, and item attributes. It is valuable to utilize information from both domains to improve the recommendation performance. In our method, we aim to simultaneously improve the recommendation accuracy in both domains.

3. Methodology

In this section, we first present the definitions and notations used in this study. Subsequently, we provide a concise overview of the framework. Finally, each module is introduced in detail.

3.1. Definitions and notations

Suppose we have two domains A and B with a shared user set $U = \{u_1, u_2, \dots, u_m\}$ (of size m) and different item sets $I^A = \{i_1^A, i_2^A, \dots, i_{n^A}^A\}$ (of size n^A), $I^B = \{i_1^B, i_2^B, \dots, i_{n^B}^B\}$ (of size n^B). A represents the source domain, and B denotes the target domain.

There are three modalities for the input data: a user-item rating matrix, user review texts, and item textual features. Let $\mathbf{R}^A \in \{0,1\}^{m \times n^A}$ and $\mathbf{R}^B \in \{0,1\}^{m \times n^B}$ represent the binary user-item interaction matrices

Table 1
Notations.

Symbols	Definitions and Notations
$*^A$, $*^B$	Domains A and B, e.g. $\mathbf{R}^{\mathbf{A}}$ represents the rating matrix in domain A
U	User set
I	Item set
m	The number of users
n	The number of items
$\mathbf{R}$	Rating matrix
G	Heterogeneous graph
$\mathbf{E}_u$	User ID embeddings
$\mathbf{E}_i$	Item ID embeddings
$\mathbf{M}$	User review embeddings
$\mathbf{T}$	Item textual embeddings
$\mathbf{H}_u$	User representations
$\mathbf{H}_i$	Item representations
$\mathbf{P}_s$	Domain-specific embeddings
$\tilde{\mathbf{P}}_s$	Augmented domain-specific embeddings
$\mathbf{P}_c$	Domain-common embeddings
$\tilde{\mathbf{P}}_c$	Augmented domain-common embeddings
$\mathbf{Q}_s$	Obfuscated domain-specific embeddings
$\tilde{\mathbf{Q}}_s$	Augmented obfuscated domain-specific embeddings
$\mathbf{Q}_c$	Obfuscated domain-common embeddings
$\tilde{\mathbf{Q}}_c$	Augmented obfuscated domain-common embeddings
$\mathbf{H}_u^*$	Comprehensive user preferences

in domains A and B, respectively. First, We aggregate the interaction data within each domain to construct two heterogeneous graphs, denoted by $G^A = (U, I^A, S^A)$ and $G^B = (U, I^B, S^B)$. These graphs serve as the foundation for learning user ID embeddings $\mathbf{E}_u^A$, $\mathbf{E}_u^B$, and item ID embeddings $\mathbf{E}_i^A$, $\mathbf{E}_i^B$ within the A and B domains, respectively. Here, S^A and S^B represent the edge sets that capture the observed user-item interactions. Let $\mathbf{M}^A$ and $\mathbf{M}^B$ denote user review embeddings and $\mathbf{T}^A$ and $\mathbf{T}^B$ indicate item textual embeddings in domains A and B.

$\mathbf{H}_u^A$ and $\mathbf{H}_u^B$ represent user representations, whereas $\mathbf{H}_i^A$ and $\mathbf{H}_i^B$ indicate item representations in domains A and B. Given user representations $\mathbf{H}_u^A$ ($\mathbf{H}_u^B$), we disentangle them into domain-specific embeddings $\mathbf{P}_s^A$ ($\mathbf{P}_s^B$) and domain-common embeddings $\mathbf{P}_c^A$ ($\mathbf{P}_c^B$). Simultaneously, feature dropout is employed to generate augmented domain-specific embeddings $\tilde{\mathbf{P}}_s^A$ ($\tilde{\mathbf{P}}_s^B$) and domain-common embeddings $\tilde{\mathbf{P}}_c^A$ ($\tilde{\mathbf{P}}_c^B$). Then, we utilize LDP to add noise to the decoupled embeddings and obtain the obfuscated embeddings $\mathbf{Q}_s^A$ ($\mathbf{Q}_s^B$) and $\mathbf{Q}_c^A$ ($\mathbf{Q}_c^B$), as well as the augmented obfuscated embeddings $\tilde{\mathbf{Q}}_s^A$ ($\tilde{\mathbf{Q}}_s^B$) and $\tilde{\mathbf{Q}}_c^A$ ($\tilde{\mathbf{Q}}_c^B$). The goal of our method is to recommend the Top-K items for all users in each domain. The mathematical notations used in this study are summarized in Table 1.

3.2. Overview of the proposed model

We propose a privacy-preserving framework with multi-modal data for CDR (P2M2-CDR). Initially, the framework employs multi-modal data to decouple more informative domain-common and domain-specific embeddings. Subsequently, it introduces LDP to prevent user privacy leakage. Fig. 2 shows an overview of the proposed framework.

This framework includes the following two key modules:

- **Multi-Modal Disentangled Encoder**: It contains two components: (1) **Multi-Modal Feature Learning**: We incorporate multi-modal information, e.g., user-item interaction matrix, review texts, and text features, to learn user and item representations. (2) **Domain Disentanglement**: We utilize MLP networks to disentangle user representations into more informative domain-common and domain-specific embeddings.
- **Privacy-preserving Decoder**: This module includes three components: (1) **Decoupled Feature Obfuscation**: We introduce Laplace noise into disentangled embeddings to ensure the privacy protection of user data. (2) **Contrastive Learning**: We introduce CL with domain-intra and domain-inter losses to regulate the separation and alignment of obfuscated decoupled embeddings. (3) **Information Fusion**: We combine obfuscated domain-common and domain-specific embeddings into final user preferences.

3.3. Multi-modal disentangled encoder

This module is designed to disentangle separate and informative domain-common and domain-specific embeddings.

3.3.1. Multi-modal feature learning

In this section, we introduce multi-modal information, including a user-item interaction matrix, user review texts, and item textual features, to learn the initial user and item representations.

ID Embeddings: Motivated by the success of GNNs that are good at modeling high-dimensional, complex relationships between users and items. We introduce LightGCN [17], a simple and lightweight graph model, to learn the user and item ID embeddings. First, we construct two heterogeneous graphs G^A and G^B to depict the user-item interaction relationships in the source domain A and target domain B, where the nodes represent the user and item entities, and the edges show the relationship (whether the user and item interact) between entities.

In this study, we use graph convolution and propagation layers in LightGCN to encode the user and item ID embeddings according to the heterogeneous graphs G^A and G^B. $\mathbf{E}_l^A$ (or $\mathbf{E}_l^B$) denote the ID embeddings in the lth layer. Specifically, the embeddings $\mathbf{E}_0^A$ (or $\mathbf{E}_0^B$) are randomly initialized. Given graph G^A, $\mathbf{E}_l^A$ can be calculated as follows:

$$\mathbf{E}_l^A = (\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})\mathbf{E}_{l-1}^A, \tag{1}$$

where $\mathbf{D}$ is the diagonal matrix and $\mathbf{A}$ is the adjacency matrix. After l propagation iterations, we generate the final user ID embedding matrix $\mathbf{E}_u^A$ and item ID embedding matrix $\mathbf{E}_i^A$ by concatenating multiple embedding matrices from $\mathbf{E}_0^A$ into $\mathbf{E}_l^A$. Similarly, we obtain the final user ID embedding matrix $\mathbf{E}_u^B$ and item ID embedding matrix $\mathbf{E}_i^B$ in domain B.

User Review Embeddings: For each user, we aggregate all review texts associated with the items they rated to create user-specific review texts. We then use a pre-trained model sentence transformer [18] and an MLP layer to generate user primitive review embeddings $\mathbf{M}^A$ (or $\mathbf{M}^B$).

Item Textual Embeddings: We concatenate the item's title, categories, and description as item textual features. These textual features are then processed using the sentence transformer model and an MLP layer to obtain the final item textual embeddings $\mathbf{T}^A$ (or $\mathbf{T}^B$).

User and Item Representations: Finally, we concatenate the user ID embeddings $\mathbf{E}_u^A$ and user review embeddings $\mathbf{M}^A$ to obtain user representations $\mathbf{H}_u^A$ in domain A:

$$\mathbf{H}_u^A = f(\mathbf{E}_u^A, \mathbf{M}^A), \tag{2}$$

where f denotes the concatenation function. Similarly, we can obtain user representations $\mathbf{H}_u^B$ in domain B.

For item representations, we concatenate item ID embeddings $\mathbf{E}_i^A$ and item textual embeddings $\mathbf{T}^A$ to learn the item representations $\mathbf{H}_i^A$:

$$\mathbf{H}_i^A = f(\mathbf{E}_i^A, \mathbf{T}^A). \tag{3}$$

Similarly, we can obtain item representations $\mathbf{H}_i^B$ in domain B.

3.3.2. Domain disentanglement

Here, we decouple the user representations $\mathbf{H}_u^A$ into more informative domain-specific embeddings $\mathbf{P}_s^A$ and domain-common embeddings $\mathbf{P}_c^A$ using two MLP layers:

$$\mathbf{P}_s^A = MLP(\mathbf{H}_u^A; \Theta_s^A); \qquad \mathbf{P}_c^A = MLP(\mathbf{H}_u^A; \Theta_c^A). \tag{4}$$

Simultaneously, it disentangles $\mathbf{H}_u^B$ into $\mathbf{P}_s^B$ and $\mathbf{P}_c^B$:

$$\mathbf{P}_s^B = MLP(\mathbf{H}_u^B; \Theta_s^B); \qquad \mathbf{P}_c^B = MLP(\mathbf{H}_u^B; \Theta_c^B), \tag{5}$$

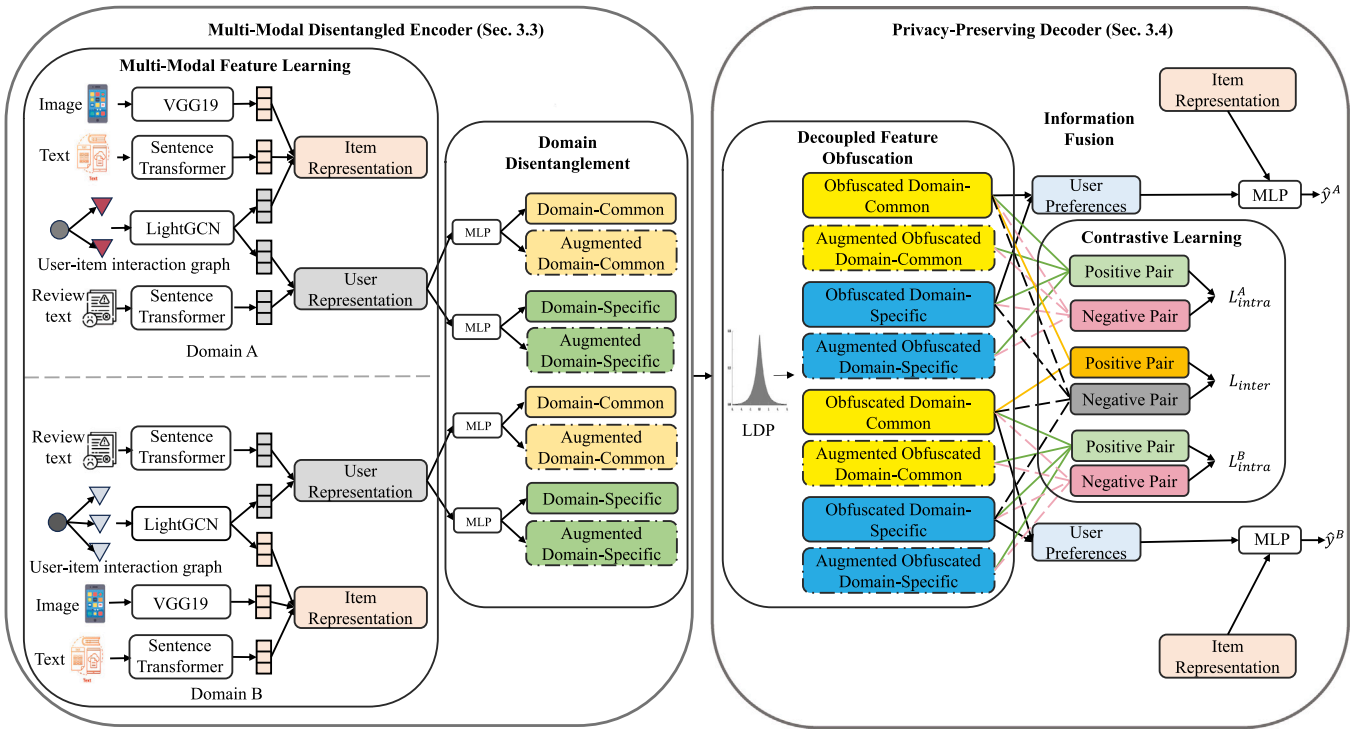


Fig. 2. The framework of P2M2-CDR. It contains two modules: (1) Multi-Modal Disentangled Encoder, which first incorporates multi-modal information, w.r.t. user-item rating matrix, review texts and textual features to learn initial user and item representations and then disentangles user representations into domain-common and domain-specific embeddings. It contains multi-modal feature learning and domain disentanglement components. (2) Privacy-Preserving Decoder, which introduces LDP to safeguard user privacy. This module includes decoupled feature obfuscation, contrastive learning, and information fusion components.

where Θ_s^A, Θ_c^A, Θ_s^B and Θ_c^B are the parameters of the respective MLP layers.

To enhance the informativeness of disentangled representations, we employ the feature dropout strategy introduced by Zhou [42] to generate augmented decoupled embeddings in both domains, denoted as $\tilde{\mathbf{P}}_s^A$, $\tilde{\mathbf{P}}_c^A$ in domain A, and $\tilde{\mathbf{P}}_s^B$, $\tilde{\mathbf{P}}_c^B$ in domain B.

3.4. Privacy-preserving decoder

In this section, we discuss three key challenges: (1) How can domain-common and domain-specific embeddings be obfuscated to protect the user-related information? (2) How can we ensure that obfuscated, disentangled embeddings do not contain redundant information within their domains, while still including shared features across domains to enhance informativeness? (3) How can we aggregate obfuscated domain-common and domain-specific embeddings to learn comprehensive and diverse user preferences? To address these issues, we introduce decoupled feature obfuscation, CL, and information fusion components.

3.4.1. Decoupled feature obfuscation

In this section, the disentangled embeddings from both the source and target domains are shared when CL is introduced to achieve alignment and separation. However, this directly carries the potential risk of inferring users' actual data [33]. To mitigate this risk and prevent the potential leakage of user's sensitive data, we employ LDP techniques to obscure the true distribution of these disentangled embeddings. Specifically, we add Laplace noise to the disentangled embeddings, which not only safeguards against external attackers intercepting the original decoupled embeddings but also prevents the other domain from inferring the user's sensitive information from these embeddings. The obfuscated disentangled embeddings can be obtained as follows:

$$\begin{aligned} \mathbf{Q}_s^A &= \mathbf{P}_s^A + La(\mu,\lambda); \quad & \mathbf{Q}_c^A &= \mathbf{P}_c^A + La(\mu,\lambda); \\ \tilde{\mathbf{Q}}_s^A &= \tilde{\mathbf{P}}_s^A + La(\mu,\lambda); \quad & \tilde{\mathbf{Q}}_c^A &= \tilde{\mathbf{P}}_c^A + La(\mu,\lambda), \end{aligned} \tag{6}$$

where μ and λ represent the mean and standard deviation of the Laplace noise, respectively. Here, we set $\mu = 0$.

Similarly, we can obtain the obfuscated domain-specific embeddings $\mathbf{Q}_s^B$, $\tilde{\mathbf{Q}}_s^B$ and domain-common embeddings $\mathbf{Q}_c^B$, $\tilde{\mathbf{Q}}_c^B$ in domain B.

3.4.2. Contrastive learning

In the last subsection, we obtain obfuscated domain-common and domain-specific features. However, we cannot ensure that these embeddings include different aspects within each domain. Moreover, we cannot ensure whether the obfuscated domain-common features in the source and target domains are indeed similar. Herein, we introduce CL-based domain-intra and domain-inter losses to address the above challenges. The core idea is to bring the positive sample pairs closer together and push the negative sample pairs farther apart.

To separate obfuscated domain-common and domain-specific embeddings, we treat variants of the same disentangled embedding as positive pairs and variants of different disentangled embeddings as negative pairs. For example, $(\mathbf{Q}_c^A, \tilde{\mathbf{Q}}_c^A)$ is a positive sample pair, $(\mathbf{Q}_c^A, \mathbf{Q}_s^A)$ and $(\mathbf{Q}_c^A, \tilde{\mathbf{Q}}_s^A)$ are negative sample pairs. The domain-intra loss is defined as follows:

$$\begin{aligned} l_{intra}^p &= exp(f(\mathbf{Q}_c^A, \tilde{\mathbf{Q}}_c^A)/\tau) + exp(f(\mathbf{Q}_s^A, \tilde{\mathbf{Q}}_s^A)/\tau); \\ l_{intra}^n &= exp(f(\mathbf{Q}_c^A, \mathbf{Q}_s^A)/\tau) + exp(f(\mathbf{Q}_c^A, \tilde{\mathbf{Q}}_s^A)/\tau) \\ &\quad + exp(f(\mathbf{Q}_s^A, \tilde{\mathbf{Q}}_c^A)/\tau) + exp(f(\tilde{\mathbf{Q}}_s^A, \tilde{\mathbf{Q}}_c^A)/\tau); \\ L_{intra}^A &= -log \frac{l_{intra}^p}{l_{intra}^p + l_{intra}^n}, \end{aligned} \tag{7}$$

where f represents the similarity function and τ is a trainable hyper-parameter of temperature. Similarly, we can obtain the domain-intra loss L_{intra}^B in the domain D^B.

For the domain-inter loss, our focus is on aligning $\mathbf{Q}_c^A$ and $\mathbf{Q}_c^B$ to ensure that they contain domain-common information while simultaneously separating $\mathbf{Q}_s^A$ and $\mathbf{Q}_s^B$ to include domain-specific information. We consider $(\mathbf{Q}_c^A, \mathbf{Q}_c^B)$ as a positive sample pair, $(\mathbf{Q}_c^A, \mathbf{Q}_s^B)$, $(\mathbf{Q}_c^B, \mathbf{Q}_s^A)$ and

$(\mathbf{Q}_s^A, \mathbf{Q}_s^B)$ as negative sample pairs. The domain-inter loss is calculated as follows:

$$\begin{aligned} l_{inter}^p &= exp(f(\mathbf{Q}_c^A, \mathbf{Q}_c^B)/\tau) \\ l_{inter}^n &= exp(f(\mathbf{Q}_c^A, \mathbf{Q}_s^B)/\tau) + exp(f(\mathbf{Q}_c^B, \mathbf{Q}_s^A)/\tau) \\ &\quad + exp(f(\mathbf{Q}_s^A, \mathbf{Q}_s^B)/\tau) \\ L_{inter} &= -log \frac{l_{inter}^p}{l_{inter}^p + l_{inter}^n}. \end{aligned} \tag{8}$$

Finally, the objective of CL is as follows:

$$L_C = L_{intra}^A + L_{intra}^B + L_{inter}. \tag{9}$$

3.4.3. Information fusion

Domain-common and domain-specific embeddings constitute two fundamental components of user preferences, and it is imperative to integrate them rationally and effectively to comprehensively capture user preferences. Therefore, we employ three fusion methods, namely element-wise sum, concatenation, and element-wise averaging, to aggregate individual embeddings into comprehensive user preferences. Following the experimental validation described in Section 4, we selected the concatenation method to generate comprehensive user preferences:

$$H_u^{A*} = g(\mathbf{Q}_c^A, \mathbf{Q}_s^A), \tag{10}$$

where g is the concatenation function. Similarly, we can obtain the comprehensive user preferences H_u^{B*} in domain B.

3.5. Model training and optimization

In this section, we feed both the user preferences H_u^{A*} (H_u^{B*}) and item representations H_i^A (H_i^B) into MLP layers to predict the probability of a user clicking on a certain item. We aim to minimize the following loss function:

$$L_{prd}^A = \sum_{r \in r^{A+} \cup r^{A-}} l(\hat{r}, r), \tag{11}$$

where l is the cross-entropy loss function. The total loss function is defined as follows:

$$L = L_{prd}^A + L_{prd}^B + \alpha L_C, \tag{12}$$

where α is the weight parameter for CL loss.

3.6. Privacy analysis

In this section, we focus on the privacy analysis of our framework P2M2-CDR. Our work aims to prevent the leakage of user privacy information, which is a significant concern in existing CDR methods.

In our approach, the data are stored within the central servers of each company (domain). These two domains communicate with each other as we introduce CL-based domain-intra and domain-inter losses to regulate the alignment and separation of disentangled embeddings.

Specifically, during the model training, we encourage the common features of these two domains to become increasingly similar, while enhancing the distinctiveness of specific features. Throughout this process, these two domains exchange disentangled embeddings, which carries the risk of inferring sensitive data from the users. For example, attackers may attempt to infer users' cross-domain behaviors or preferences from the information shared across domains. Consequently, we employ the LDP method to conceal disentangled embeddings. This is achieved by introducing noise that adheres to the Laplace distribution.

Within LDP, a higher standard deviation (λ), which controls noise strength, can safeguard sensitive data more effectively, thereby reducing the risk of data leakage. However, this may lead to a reduction in the recommendation performance. Hence, it is essential to set an appropriate value for λ to strike a balance between recommendation accuracy and privacy protection.

Table 2
Statistic of the datasets for six CDR tasks.

Tasks	Datasets	#Users	#Items	#ratings	Density
Task 1	Phone	22,998	38,800	186,365	0.0208%
	Elec	22,998	77,187	364,029	0.0205%
Task 2	Phone	5,902	18,635	56,112	0.051%
	Sport	5,902	29,180	67,276	0.039%
Task 3	Sport	12,965	46,868	134,676	0.022%
	Cloth	12,965	62,343	134,467	0.017%
Task 4	Elec	19,754	69,362	279,282	0.020%
	Cloth	19,754	77,003	179,735	0.012%
Task 5	Book	1688	8660	103,994	0.711%
	Movie	1688	20,558	889,970	2.56%
Task 6	Movie	1176	20,543	699,917	2.90%
	Music	1176	7146	76,051	0.905%

4. Experiments

In this section, to evaluate the performance of our proposed model, P2M2-CDR, we conduct a series of comprehensive experiments on the widely used Amazon and Douban datasets to answer the following questions:

- RQ1: Does our model achieve superior performance compared to other state-of-the-art baseline methods?
- RQ2: How do different components, such as CL, decoupled feature obfuscation, domain-specific information, domain-common information, and multi-modal information, influence the outcomes of our model?
- RQ3: Are the embeddings we have acquired genuinely disentangled?
- RQ4: How does our model's performance vary with different hyperparameters?
- RQ5: How to maintain a balance between privacy protection and model performance?

4.1. Experimental settings

4.1.1. Datasets

We conduct comprehensive experiments on four real-world benchmark subsets from the Amazon dataset[1]: Cell Phones and Accessories (Phone), Electronics (Elec), Sports and Outdoors (Sport), and Clothing, Shoes and Jewelry (Cloth), as well as three subsets from the Douban dataset[2]: Book, Movie, and Music. These datasets are widely used in CDR [9,11,37]. We combine them into six CDR tasks. Table 2 presents the basic statistics for these datasets. For these datasets, we transform the explicit ratings into implicit feedback. Specifically, we discretize the ratings into binary values of 0 and 1 to indicate whether the user has interacted with an item. For each observed user-item interaction, we randomly select an item with which the user had not previously interacted as a negative sample. To ensure data quality and alleviate sparsity issues, filtering criteria are applied to remove records with fewer than five interactions between all users and items across both domains. The multi-modal features of these datasets include user-item interaction history, text features, and review texts.

4.1.2. Evaluation metrics

Motivated by BiTGCF [9] and DIDA-CDR [14], we use the leave-one-out method to evaluate the model's performance. For each user, we randomly select one sample to create the test set, and the remaining samples form the training set. Following NeuMF [23], for each test

[1] https://cseweb.ucsd.edu/jmcauley/datasets/amazon/links.html.
[2] https://www.dropbox.com/s/u2ejjezjk08lz1o/Douban.tar.gz?e=2&dl=0.

user, we randomly select 99 items that the user has not interacted with as negative samples and the ground-truth user-item interaction as the positive sample. We then employ the P2M2-CDR model to predict scores for these 100 candidate items to perform ranking. The evaluation of recommendation performance relies on five metrics: Hit Ratio (HR), Normalized Discounted Cumulative Gain (NDCG), Mean Reciprocal Rank (MRR), Precision (Prec), and F1-score (F1) which are widely used in CDR methods [7,27,43].

4.1.3. Baseline methods

To verify the effectiveness of our model, we compare the performance of P2M2-CDR with three sets of representative baseline methods: Single-Domain Recommendation, Cross-Domain Recommendation and Privacy-Preserving CDR. Each set of baseline methods represents a distinct category or paradigm in the field. They have been extensively studied and serve as benchmarks for evaluating the performance of CDR methods [14,27,37,38,44]. More importantly, they leverage technologies and methodologies that are similar to those employed in the proposed model. By including these baseline methods in our comparison, we ensure a comprehensive evaluation against existing state-of-the-art approaches, enabling meaningful insights into the effectiveness and applicability of the proposed model.

Single-Domain Recommendation

- **NeuMF** [23] combines collaborative filtering and neural network techniques to capture user-item interactions and make accurate predictions.
- **LightGCN** [17] is a simple graph convolutional network (GCN) model that directly propagates user and item embeddings through the user-item interaction graph without introducing complex operations or auxiliary information.

Cross-Domain Recommendation

- **PTUPCDR** [45] is a framework focusing on personalized user preference transfer via a meta-network.
- **DDTCDR** [7] is a method that aims to learn an orthogonal mapping function to transfer user preferences across domains and provide recommendations for both domains. In addition to user-item rating information, it also utilizes user and item features to improve recommendation accuracy.
- **DCCDR** [12] is a framework that disentangles domain-invariant and domain-specific representations and then uses GNN to learn high-order relationships to enrich these representations.
- **DisenCDR** [11] focuses on disentangling user preferences into domain-specific and domain-shared information and transferring the domain-shared knowledge across domains.
- **BiTGCF** [9] is a method that conducts bidirectional high-order information transfer to enhance the performance of graph collaborative filtering-based CDR.

Privacy-Preserving CDR

- **PriCDR** [37] is a privacy-preserving CDR framework that utilizes LDP technology to publish the rating matrix in the source domain and then transfers the published matrix to the target domain.
- **P2FCDR** [38] is a privacy-preserving federated framework that learns an orthogonal mapping matrix to transform embeddings across domains and apply the LDP technique on the transformed embeddings to protect user privacy.

4.1.4. Parameter settings

We implement the P2M2-CDR model using Python with the Pytorch framework, and all baseline methods are conducted based on the GitHub source code and carefully adjusted the hyperparameters. The optimal hyperparameters are obtained by optimizing the loss function (12) using the Adam optimizer with a learning rate of 0.001. For the disentangled encoder, we use a two-layer fully connected network with dimensions of 128 and 256, and we obtain the disentangled embeddings with a dimension of 256. Considering the trade-off between recommendation performance and privacy protection, we set λ to 0.01. Simultaneously, the weight of CL loss is set to 0.005. We set the batch size to 512 and the number of epochs to 200. To prevent overfitting, batch normalization, dropout, and early stopping techniques are applied.

4.2. Performance evaluation (RQ1)

We evaluate the performance of P2M2-CDR and the baselines using commonly used evaluation metrics, w.r.t. HR@{5, 10}, NDCG@{5, 10}, MRR@{5, 10}, Prec@{5, 10}, and F1@{5, 10}, on six CDR tasks. The results are shown in Tables 3 and 4. Based on these results, we observe that:

- Our model, P2M2-CDR, surpasses other baselines and achieves superior performance, with average improvements of 11.84%, 10.22%, 8.10%, 1.18%, and 2.15% over the best baseline in terms of HR@10, NDCG@10, MRR@10, Prec@10, and F1@10, respectively. Additionally, despite introducing noise to protect user privacy, our method still outperforms other CDR methods. These results indicate that the proposed P2M2-CDR can enhance recommendation performance in both the source and target domains simultaneously while safeguarding user privacy.
- P2M2-CDR performs better on the Amazon dataset than on the Douban dataset, despite the greater density of Douban. This improved performance may be because the multi-modal features in the Amazon dataset are better suited to addressing the data-sparsity problem, thereby improving the model performance.
- P2M2-CDR outperforms the privacy-preserving CDR methods PriCDR and P2FCDR, both of which use LDP technology. This demonstrates that disentangling domain-common and domain-specific embeddings with multi-modal information can enhance model performance.
- Compared with other disentanglement-based CDR approaches such as DisenCDR and DCCDR, P2M2-CDR achieves the best performance, which indicates that multi-modal features play an important role in disentangling informative embeddings.
- CDR methods outperform single-domain recommendation models, particularly in domains with sparse data, such as DisenCDR vs. LightGCN and P2M2-CDR vs. LightGCN. This demonstrates that CDR methods are effective in addressing the data-sparsity problem.
- Dual-target CDR methods perform better than single-target CDR approaches, such as P2M2-CDR vs. PTUPCDR and F2FCDR vs. PriCDR. This is because dual-target CDR can simultaneously improve the performance of both source and target domains.
- GNN-based methods outperform non-graph methods, such as LightGCN vs. NeuMF and DCCDR vs. DisenCDR. This demonstrates that incorporating high-order neighbor information can improve model accuracy.
- Integrating user and item text features, such as DDTCDT vs. BiTGCF, can improve model performance. In addition, multi-modal features play an important role in improving the recommendation accuracy, such as our model P2M2-CDR vs. DDTCDR.

4.3. Ablation studies (RQ2)

To assess the effectiveness of each component in P2M2-CDR, we conducted ablation experiments on six CDR tasks. We created seven variants of P2M2-CDR by removing specific components.

- w/o rev: Removes user review texts when learning comprehensive user representations.

Table 3
Experimental results on tasks 1–3. The best performance is in bold, and the second best is underlined.

Datasets	Metrics		Single domain methods		CDR methods					PPCDR methods		Ours	
			NeuMF	LightGCN	PTUPCDR	BiTGCF	DisenCDR	DDTCDR	DCCDR	PriCDR	P2FCDR	P2M2-CDR	Imp
Phone	N = 5	HR	0.3307	0.3376	0.3405	0.4140	0.4189	0.5079	0.5347	0.4027	0.5135	**0.6348**	10.01%
		NDCG	0.2523	0.2603	0.2411	0.3008	0.3105	0.3338	0.3561	0.2719	0.3671	**0.4640**	9.69%
		MRR	0.2264	0.2296	0.2089	0.2566	0.2668	0.2765	0.3028	0.2136	0.3193	**0.4077**	8.84%
		Prec	0.0661	0.0675	0.0681	0.0828	0.0838	0.1016	0.1069	0.0805	0.1027	**0.1270**	2.43%
		F1	0.1102	0.1125	0.1135	0.1380	0.1396	0.1693	0.1782	0.1342	0.1712	**0.2116**	3.34%
	N = 10	HR	0.5093	0.5124	0.5289	0.5360	0.5423	0.7063	0.7234	0.6321	0.7156	**0.8157**	9.23%
		NDCG	0.3357	0.3478	0.3564	0.3568	0.3751	0.4228	0.4389	0.4028	0.4414	**0.5226**	8.12%
		MRR	0.3015	0.3042	0.3088	0.3107	0.3258	0.3239	0.3470	0.3572	0.3825	**0.4319**	4.94%
		Prec	0.0509	0.0512	0.0529	0.0536	0.0542	0.0706	0.0723	0.0632	0.0716	**0.0816**	0.93%
		F1	0.0926	0.0932	0.0962	0.0975	0.0986	0.1284	0.1315	0.1149	0.1301	**0.1483**	1.68%
Elec	N = 5	HR	0.2214	0.2305	0.2942	0.3009	0.3347	0.4657	0.4968	0.3046	0.4378	**0.5842**	8.74%
		NDCG	0.1487	0.1549	0.1640	0.1905	0.2371	0.3037	0.3401	0.1726	0.2506	**0.4292**	8.91%
		MRR	0.1247	0.1298	0.1314	0.1539	0.2038	0.2506	0.2963	0.1430	0.2084	**0.3780**	8.17%
		Prec	0.0443	0.0461	0.0588	0.0602	0.0669	0.0931	0.0994	0.0609	0.0876	**0.1168**	1.75%
		F1	0.0738	0.0768	0.0981	0.1003	0.1116	0.1552	0.1656	0.1015	0.1459	**0.1947**	2.91%
	N = 10	HR	0.4412	0.4498	0.4813	0.4971	0.5903	0.6539	0.6612	0.5515	0.6478	**0.7482**	8.70%
		NDCG	0.2848	0.2901	0.2971	0.2967	0.3935	0.4023	0.4110	0.3688	0.3967	**0.4825**	7.15%
		MRR	0.2474	0.2550	0.2614	0.2667	0.3108	0.3104	0.3301	0.3153	0.3061	**0.4001**	7.00%
		Prec	0.0441	0.0450	0.0481	0.0497	0.0590	0.0654	0.0661	0.0552	0.0648	**0.0748**	0.87%
		F1	0.0802	0.0818	0.0875	0.0904	0.1073	0.1189	0.1202	0.1003	0.1178	**0.1360**	1.58%
Phone	N = 5	HR	0.2753	0.2834	0.3501	0.4034	0.4011	0.4628	0.5023	0.3655	0.4128	**0.7047**	20.24%
		NDCG	0.2006	0.2157	0.2482	0.3053	0.2509	0.3044	0.3267	0.2981	0.2874	**0.4914**	16.47%
		MRR	0.1759	0.1794	0.2152	0.2595	0.2310	0.2525	0.2676	0.2460	0.2561	**0.4211**	15.35%
		Prec	0.0551	0.0567	0.0700	0.0807	0.0802	0.0926	0.1005	0.0731	0.0826	**0.1409**	4.05%
		F1	0.0918	0.0945	0.1167	0.1345	0.1337	0.1543	0.1674	0.1218	0.1376	**0.2349**	6.75%
	N = 10	HR	0.3531	0.3624	0.4857	0.5076	0.5314	0.5504	0.6606	0.4958	0.6302	**0.8909**	23.03%
		NDCG	0.2296	0.2307	0.2534	0.2740	0.3476	0.3601	0.3887	0.3318	0.3838	**0.5524**	16.37%
		MRR	0.1861	0.1934	0.1657	0.2155	0.2615	0.2824	0.3256	0.3167	0.3267	**0.4468**	12.01%
		Prec	0.0353	0.0362	0.0486	0.0508	0.0531	0.0550	0.0661	0.0496	0.0630	**0.0891**	2.30%
		F1	0.0642	0.0659	0.0883	0.0923	0.0966	0.1001	0.1201	0.0901	0.1146	**0.1620**	4.19%
Sport	N = 5	HR	0.1971	0.2108	0.2500	0.3092	0.3058	0.3449	0.4983	0.3440	0.4378	**0.7687**	27.04%
		NDCG	0.1414	0.1548	0.1771	0.1792	0.1873	0.2133	0.2879	0.2446	0.2954	**0.5476**	25.22%
		MRR	0.1231	0.1305	0.1438	0.1430	0.1526	0.1705	0.2451	0.2118	0.2538	**0.4743**	22.05%
		Prec	0.0394	0.0422	0.0500	0.0618	0.0612	0.0690	0.0997	0.0688	0.0876	**0.1537**	5.41%
		F1	0.0657	0.0703	0.0833	0.1031	0.1019	0.1150	0.1661	0.1147	0.1459	**0.2562**	9.01%
	N = 10	HR	0.3053	0.3214	0.4873	0.5290	0.5323	0.5672	0.6325	0.5315	0.6270	**0.9075**	27.50%
		NDCG	0.1808	0.1956	0.2228	0.2642	0.3315	0.3521	0.3894	0.3789	0.3801	**0.5933**	20.39%
		MRR	0.1515	0.1663	0.1461	0.1721	0.1894	0.2023	0.3251	0.3061	0.3247	**0.4937**	16.86%
		Prec	0.0305	0.0321	0.0487	0.0529	0.0532	0.0567	0.0633	0.0532	0.0627	**0.0908**	2.75%
		F1	0.0555	0.0584	0.0886	0.0962	0.0968	0.1031	0.1150	0.0966	0.1140	**0.1650**	5.00%
Sport	N = 5	HR	0.1834	0.1968	0.2448	0.4046	0.4136	0.4349	0.5034	0.3522	0.5012	**0.6025**	9.91%
		NDCG	0.1278	0.1346	0.1434	0.3020	0.3201	0.3042	0.3559	0.2840	0.3427	**0.4111**	5.52%
		MRR	0.1097	0.1165	0.1105	0.2681	0.2846	0.2618	0.2940	0.2615	0.2938	**0.3481**	5.41%
		Prec	0.0367	0.0394	0.0490	0.0809	0.0827	0.0870	0.1007	0.0704	0.1002	**0.1205**	1.98%
		F1	0.0611	0.0656	0.0816	0.1349	0.1379	0.1450	0.1678	0.1174	0.1671	**0.2008**	3.30%
	N = 10	HR	0.3459	0.3562	0.4825	0.5409	0.5623	0.5912	0.7342	0.5461	0.7196	**0.8033**	6.91%
		NDCG	0.2131	0.2257	0.2481	0.3069	0.3351	0.3652	0.4124	0.3186	0.3976	**0.4762**	6.38%
		MRR	0.1814	0.1894	0.2100	0.2455	0.2702	0.2745	0.3021	0.2795	0.3002	**0.3750**	7.29%
		Prec	0.0346	0.0356	0.0483	0.0541	0.0562	0.0591	0.0734	0.0546	0.0720	**0.0803**	0.69%
		F1	0.0629	0.0648	0.0877	0.0983	0.1022	0.1075	0.1335	0.0993	0.1308	**0.1461**	1.26%
Cloth	N = 5	HR	0.1585	0.1675	0.2420	0.3863	0.3958	0.3559	0.4349	0.3583	0.4339	**0.5382**	10.33%
		NDCG	0.1097	0.1149	0.1626	0.2438	0.2601	0.2304	0.2976	0.2404	0.2851	**0.3728**	7.52%
		MRR	0.0936	0.1105	0.1305	0.2070	0.2149	0.1896	0.2457	0.2079	0.2403	**0.3185**	7.28%
		Prec	0.0317	0.0335	0.0484	0.0773	0.0792	0.0712	0.0870	0.0717	0.0868	**0.1076**	2.07%
		F1	0.0528	0.0558	0.0807	0.1288	0.1319	0.1186	0.1450	0.1194	0.1446	**0.1794**	3.44%
	N = 10	HR	0.2448	0.2592	0.4821	0.5491	0.5812	0.6078	0.6946	0.5330	0.6903	**0.7318**	3.72%
		NDCG	0.1523	0.1645	0.2191	0.3268	0.3301	0.3554	0.4023	0.3237	0.3778	**0.4354**	3.31%
		MRR	0.1329	0.1450	0.1693	0.2588	0.2630	0.2716	0.2947	0.2868	0.3092	**0.3443**	3.51%
		Prec	0.0245	0.0259	0.0482	0.0549	0.0581	0.0608	0.0695	0.0533	0.0690	**0.0732**	0.37%
		F1	0.0445	0.0471	0.0877	0.0998	0.1057	0.1105	0.1263	0.0969	0.1255	**0.1331**	0.68%

- w/o txt: Removes item textual features when learning comprehensive item representations.
- w/o com: Removes obfuscated domain-common features for recommendation.
- w/o spe: Removes obfuscated domain-specific features for recommendation.
- w/o intra: Eliminates the domain-intra CL loss.
- w/o inter: Removes the domain-inter CL loss.
- w/o obf: Deletes the decoupled feature obfuscation module.

The results of the ablation studies are presented in Table 5. Based on these results, we make the following observations:

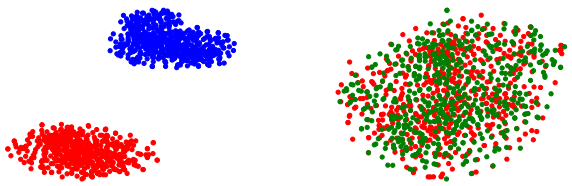
- Without user review texts, the performance of 'w/o rev' declines by an average of 24.66%, 15.98%, 13.30%, 2.47%, and 4.48% in

Table 4
Experimental results on tasks 4–6. The best performance is in bold, and the second best is underlined.

Datasets	Metrics		Single domain methods		CDR methods					PPCDR methods		Ours	
			NeuMF	LightGCN	PTUPCDR	BiTGCF	DisenCDR	DDTCDR	DCCDR	PriCDR	P2FCDR	P2M2-CDR	Imp
Elec	N = 5	HR	0.2803	0.2905	0.2949	0.4390	0.4435	0.4292	0.5358	0.4456	0.5145	**0.6094**	7.36%
		NDCG	0.2125	0.2234	0.2151	0.3367	0.3398	0.2932	0.3397	0.2249	0.3209	**0.4338**	9.41%
		MRR	0.1900	0.2056	0.1928	0.3028	0.3046	0.2487	0.2805	0.1620	0.2764	**0.3759**	9.54%
		Prec	0.0561	0.0581	0.0590	0.0878	0.0887	0.0858	0.1072	0.0891	0.1029	**0.1219**	1.47%
		F1	0.0934	0.0968	0.0983	0.1463	0.1478	0.1431	0.1786	0.1485	0.1715	**0.2031**	2.45%
	N = 10	HR	0.3459	0.4120	0.4796	0.6366	0.6414	0.6514	0.7069	0.6104	0.6845	**0.7889**	8.20%
		NDCG	0.2131	0.2635	0.2192	0.3429	0.3517	0.3552	0.4112	0.4134	0.3672	**0.4920**	7.86%
		MRR	0.1714	0.1903	0.1621	0.2868	0.2904	0.3060	0.3357	0.3467	0.3027	**0.4001**	5.34%
		Prec	0.0346	0.0412	0.0480	0.0637	0.0641	0.0651	0.0707	0.0610	0.0685	**0.0789**	0.82%
		F1	0.0629	0.0749	0.0872	0.1157	0.1166	0.1184	0.1285	0.1110	0.1245	**0.1434**	1.49%
Cloth	N = 5	HR	0.1545	0.1598	0.2411	0.3368	0.3425	0.4023	0.4506	0.3137	0.4219	**0.6168**	16.62%
		NDCG	0.1103	0.1245	0.1424	0.2015	0.2146	0.2021	0.2468	0.1746	0.2173	**0.4209**	17.41%
		MRR	0.0958	0.1027	0.1104	0.1469	0.1501	0.1639	0.2076	0.1239	0.1567	**0.3564**	14.88%
		Prec	0.0309	0.0320	0.0482	0.0674	0.0685	0.0805	0.0901	0.0627	0.0844	**0.1234**	3.32%
		F1	0.0515	0.0533	0.0804	0.1123	0.1142	0.1341	0.1502	0.1046	0.1406	**0.2056**	5.54%
	N = 10	HR	0.2448	0.2612	0.4828	0.5975	0.6070	0.6198	0.6578	0.5749	0.6503	**0.8206**	16.28%
		NDCG	0.1523	0.1478	0.2188	0.3479	0.3467	0.3592	0.3612	0.3412	0.3579	**0.4872**	12.60%
		MRR	0.1260	0.1302	0.1395	0.2010	0.2067	0.2489	0.2587	0.2485	0.2683	**0.3840**	11.57%
		Prec	0.0245	0.0261	0.0483	0.0598	0.0607	0.0620	0.0658	0.0575	0.0650	**0.0821**	1.63%
		F1	0.0445	0.0475	0.0878	0.1086	0.1104	0.1127	0.1196	0.1045	0.1182	**0.1492**	2.96%
Book	N = 5	HR	0.2227	0.2345	0.3549	0.4070	0.4151	0.4193	0.4531	0.2346	0.4328	**0.5646**	11.15%
		NDCG	0.1548	0.1657	0.2390	0.2918	0.2837	0.2687	0.3094	0.1644	0.2268	**0.3981**	8.87%
		MRR	0.1323	0.1358	0.1748	0.2537	0.2278	0.2161	0.2448	0.1414	0.1736	**0.3431**	9.83%
		Prec	0.0445	0.0469	0.0710	0.0814	0.0830	0.0839	0.0906	0.0469	0.0866	**0.1129**	2.23%
		F1	0.0742	0.0782	0.1183	0.1357	0.1384	0.1398	0.1510	0.0782	0.1443	**0.1882**	3.72%
	N = 10	HR	0.2885	0.2907	0.5023	0.5332	0.5524	0.5969	0.6358	0.3270	0.5873	**0.7186**	8.28%
		NDCG	0.1759	0.1846	0.2201	0.3324	0.3368	0.3448	0.3649	0.1916	0.3562	**0.4482**	8.33%
		MRR	0.1409	0.1398	0.1606	0.2704	0.3036	0.3071	0.3126	0.1500	0.3021	**0.3640**	5.14%
		Prec	0.0289	0.0291	0.0502	0.0533	0.0552	0.0597	0.0636	0.0327	0.0587	**0.0719**	0.83%
		F1	0.0525	0.0529	0.0913	0.0969	0.1004	0.1085	0.1156	0.0595	0.1068	**0.1307**	1.51%
Movie	N = 5	HR	0.2603	0.2748	0.3362	0.4064	0.4098	0.3943	0.5109	0.3312	0.4014	**0.6155**	10.46%
		NDCG	0.1839	0.1920	0.2530	0.2925	0.3048	0.3284	0.3443	0.2667	0.3338	**0.4357**	9.14%
		MRR	0.1456	0.1482	0.2048	0.2549	0.2604	0.2840	0.3038	0.2159	0.2937	**0.3763**	7.25%
		Prec	0.0521	0.0550	0.0672	0.0813	0.0820	0.0789	0.1022	0.0662	0.0803	**0.1231**	2.09%
		F1	0.0868	0.0916	0.1121	0.1355	0.1366	0.1314	0.1703	0.1104	0.1338	**0.2052**	3.49%
	N = 10	HR	0.3120	0.3169	0.5039	0.5379	0.5539	0.6242	0.6630	0.5207	0.6458	**0.7613**	9.83%
		NDCG	0.2013	0.2174	0.2235	0.3349	0.3520	0.3742	0.3952	0.2891	0.4035	**0.4825**	7.90%
		MRR	0.1647	0.1705	0.1806	0.2723	0.3127	0.3441	0.3467	0.2480	0.3429	**0.3955**	4.88%
		Prec	0.0312	0.0317	0.0504	0.0538	0.0554	0.0624	0.0663	0.0521	0.0646	**0.0761**	0.98%
		F1	0.0567	0.0576	0.0916	0.0978	0.1007	0.1135	0.1205	0.0947	0.1174	**0.1384**	1.79%
Movie	N = 5	HR	0.2311	0.2378	0.2647	0.4088	0.4123	0.4465	0.4927	0.4677	0.4835	**0.5468**	5.41%
		NDCG	0.1297	0.1328	0.1516	0.2610	0.2749	0.2995	0.3409	0.3096	0.2635	**0.3800**	3.91%
		MRR	0.1063	0.1135	0.1151	0.2121	0.2236	0.2180	0.2938	0.2577	0.2259	**0.3253**	3.15%
		Prec	0.0462	0.0476	0.0529	0.0818	0.0825	0.0893	0.0985	0.0935	0.0967	**0.1094**	1.08%
		F1	0.0770	0.0793	0.0882	0.1363	0.1374	0.1488	0.1642	0.1559	0.1612	**0.1823**	1.80%
	N = 10	HR	0.3046	0.3087	0.5081	0.5670	0.5739	0.6086	0.6509	0.5757	0.6421	**0.7168**	6.59%
		NDCG	0.2057	0.2105	0.2893	0.3123	0.3261	0.3495	0.3487	0.2994	0.3309	**0.4349**	8.54%
		MRR	0.1794	0.1847	0.2265	0.2334	0.2682	0.2891	0.2902	0.2153	0.2871	**0.3478**	5.76%
		Prec	0.0305	0.0309	0.0508	0.0567	0.0574	0.0609	0.0651	0.0576	0.0642	**0.0717**	0.66%
		F1	0.0554	0.0561	0.0924	0.1031	0.1043	0.1107	0.1183	0.1047	0.1167	**0.1303**	1.20%
Music	N = 5	HR	0.1488	0.1537	0.2570	0.3019	0.3219	0.3977	0.4862	0.2446	0.4204	**0.6760**	18.98%
		NDCG	0.0985	0.1034	0.1481	0.2114	0.2301	0.2607	0.3516	0.1829	0.3130	**0.5258**	17.42%
		MRR	0.0821	0.0925	0.1127	0.1816	0.2019	0.2158	0.3174	0.1628	0.2562	**0.4759**	15.85%
		Prec	0.0298	0.0307	0.0514	0.0604	0.0644	0.0795	0.0972	0.0489	0.0841	**0.1352**	3.80%
		F1	0.0496	0.0512	0.0857	0.1006	0.1073	0.1326	0.1621	0.0815	0.1401	**0.2253**	6.33%
	N = 10	HR	0.2338	0.2396	0.5030	0.6058	0.6134	0.6299	0.6572	0.4449	0.6073	**0.7951**	13.79%
		NDCG	0.1260	0.1342	0.2247	0.3210	0.3198	0.3353	0.4073	0.2195	0.4120	**0.5644**	15.71%
		MRR	0.0934	0.1035	0.1427	0.2346	0.2405	0.2464	0.3627	0.1817	0.3519	**0.4919**	12.92%
		Prec	0.0234	0.0240	0.0503	0.0606	0.0613	0.0630	0.0657	0.0445	0.0607	**0.0795**	1.38%
		F1	0.0425	0.0436	0.0915	0.1101	0.1115	0.1145	0.1195	0.0809	0.1104	**0.1446**	2.51%

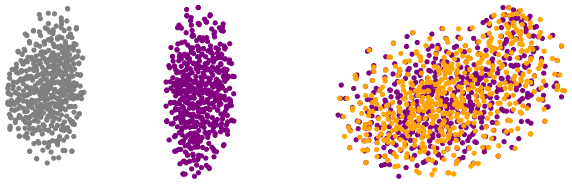
terms of HR@10, NDCG@10, MRR@10, Prec@10, and F1@10, respectively, compared to P2M2-CDR. This highlights the significance of user review texts in disentangling domain-common and domain-specific features and in modeling comprehensive user representations.

- The model 'w/o txt', which eliminates the item textual features, experiences an average drop of 28.87%, 18.05%, 14.63%, 2.89%, and 5.25% in terms of HR@10, NDCG@10, MRR@10, Prec@10, and F1@10 respectively. These results demonstrate the significance of incorporating textual features of items in the process of learning comprehensive item representations.
- Comparing P2M2-CDR with 'w/o com' and 'w/o spe', we can see that obfuscated disentangled domain-common and domain-specific features play a vital role in predicting user preferences.



(a) (b)

Fig. 3. Visualization of user obfuscated disentangled embeddings in Task 2: **Phone&Sport**. (a) Red points represent obfuscated domain-common embeddings and blue points indicate obfuscated domain-specific embeddings in the source domain (**Phone**); (b) Red points represent obfuscated domain-common embeddings in the source domain (**Phone**) and green points show the obfuscated domain-common embeddings in the target domain (**Sport**).



(a) (b)

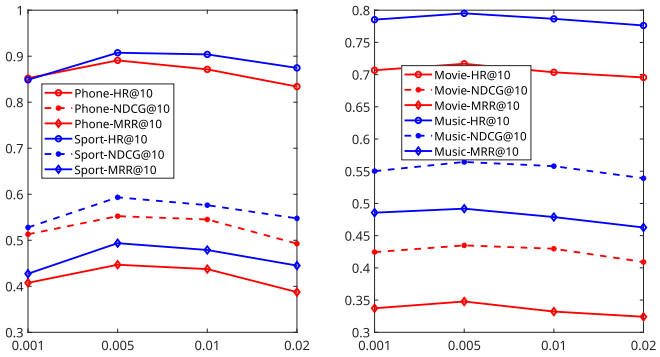
Fig. 4. Visualization of user obfuscated disentangled embeddings in Task 6: **Movie&Music**. (a) Purple points represent obfuscated domain-common embeddings and gray points indicate obfuscated domain-specific embeddings in the source domain (**Movie**); (b) Purple points represent obfuscated domain-common embeddings in the source domain (**Movie**) and orange points show the obfuscated domain-common embeddings in the target domain (**Music**).

- The inferior performance of models 'w/o intra' and 'w/o inter' further shows the significant contributions of both domain-intra and domain-inter CL losses to the final outcome.
- It is worth noting that the model 'w/o obf', without the addition of noise, outperforms P2M2-CDR. However, this approach may pose a privacy threat by potentially exposing user data. We should carefully balance the trade-off between ensuring privacy protection and achieving optimal model performance.

4.4. Visualization of obfuscated disentangled embeddings (RQ3)

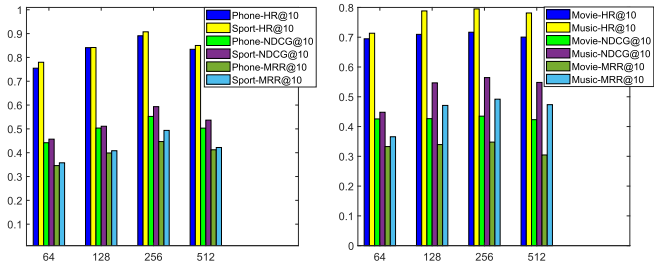
To determine whether the obfuscated domain-common and domain-specific features are distinct and contain diverse information, we visualize these embeddings using t-SNE [46]. This is a data visualization technique that projects high-dimensional data onto a lower-dimensional space. We randomly select 1000 users in both domains for CDR Tasks 2 (Phone&Sport) and 6 (Movie&Music). The visualization results are shown in Figs. 3(a) and 4(a). Furthermore, to assess whether the proposed model has effectively acquired domain-common knowledge from both source and target domains through shared users, we visualize the obfuscated domain-common features in both domains in Figs. 3(b) and 4(b).

A distinct separation is observed between the domain-common and domain-specific embeddings in Figs. 3(a) and 4(a). This clear separation confirms that the proposed model can effectively disentangle user representations, ensuring that the obfuscated disentangled embeddings do not contain redundant information. In Figs. 3(b) and 4(b), it is evident that these features are similar, which demonstrates that the obfuscated domain-common embeddings contain shared information between the two domains.



(a) Phone&Sport (b) Movie&Music

Fig. 5. Performance of different weight parameter α.



(a) Phone&Sport (b) Movie&Music

Fig. 6. Performance of different decoupled embedding dimensions.

4.5. Impact of hyperparameters (RQ4)

In this section, we evaluate the model's performance in terms of HR@10, NDCG@10, and MRR@10 across various settings for four crucial parameters: the weight parameter α for CL losses, the decoupling embedding dimension, the recommendation list length K, and the information aggregation method for disentangled embeddings in Tasks 2 and 6.

4.5.1. Impact of α

To validate the effect of the parameter α, we test the performance of the model with different values of α [0.001, 0.005, 0.01, 0.02] and present the results in Fig. 5. We observe that our model achieves its peak performance when α is set as 0.005. As α increases, the HR@10, NDCG@10, and MRR@10 metrics initially rise, peak at $\alpha = 0.005$, and then decline. This suggests that an α value of 0.005 achieves an optimal balance between the CL and predictive losses.

4.5.2. Impact of decoupled embedding dimension

In this section, we evaluate the model's performance using different decoupled embedding sizes [64, 128, 256, 512], as shown in Fig. 6. We observe that the optimal performance is attained when the decoupled embedding dimension is set to 256. Increasing the decoupled embedding dimension improves the model's effectiveness. However, it is important to note that too large decoupled embedding dimensions may lead to overfitting. Thus, choosing an appropriate decoupled embedding dimension is crucial to balance model complexity and performance.

4.5.3. Impact of recommendation list length K

To evaluate the impact of the recommendation list length on performance, we test the performance of P2M2-CDR using different values

Table 5
Ablation studies on all tasks.

Tasks	Datasets	Metrics	Variants							
			w/o rev	w/o txt	w/o com	w/o spe	w/o intra	w/o inter	w/o obf	P2M2-CDR
Task1	Phone	HR	0.5712	0.5643	0.8036	0.8034	0.7003	0.7606	0.8533	0.8157
		NDCG	0.3634	0.3577	0.4944	0.5000	0.4073	0.4606	0.5524	0.5226
		MRR	0.2994	0.2937	0.3880	0.4053	0.3091	0.3643	0.4588	0.4319
		Prec	0.0571	0.0564	0.0804	0.0803	0.0700	0.0761	0.0853	0.0816
		F1	0.1038	0.1026	0.1462	0.1460	0.1273	0.1384	0.1551	0.1484
	Elec	HR	0.5651	0.4899	0.7051	0.7095	0.6892	0.6024	0.7597	0.7482
		NDCG	0.3702	0.3276	0.4468	0.4060	0.4100	0.3579	0.4901	0.4825
		MRR	0.3099	0.2774	0.3602	0.3711	0.3234	0.3092	0.4066	0.4001
		Prec	0.0565	0.0490	0.0705	0.0710	0.0690	0.0602	0.0760	0.0748
		F1	0.1027	0.0891	0.1282	0.1291	0.1254	0.1095	0.1382	0.1360
Task2	Phone	HR	0.4792	0.4905	0.8790	0.8585	0.7066	0.7158	0.9044	0.8909
		NDCG	0.2878	0.2971	0.5443	0.5247	0.4327	0.4508	0.5710	0.5524
		MRR	0.2293	0.2376	0.4401	0.4210	0.3709	0.4061	0.4663	0.4468
		Prec	0.0479	0.0491	0.0879	0.0859	0.0707	0.0716	0.0904	0.0891
		F1	0.0871	0.0893	0.1598	0.1562	0.1285	0.1302	0.1644	0.1620
	Sport	HR	0.4121	0.3966	0.8897	0.8812	0.7680	0.7756	0.9083	0.9075
		NDCG	0.2248	0.2222	0.5807	0.5681	0.4454	0.4547	0.5996	0.5933
		MRR	0.1680	0.1690	0.4834	0.4696	0.4041	0.4093	0.4995	0.4937
		Prec	0.0412	0.0397	0.0890	0.0881	0.0768	0.0776	0.0908	0.0908
		F1	0.0749	0.0722	0.1618	0.1602	0.1396	0.1411	0.1651	0.1651
Task3	Sport	HR	0.4121	0.3966	0.8897	0.8812	0.7680	0.7756	0.9083	0.9075
		NDCG	0.2248	0.2222	0.5807	0.5681	0.4454	0.4547	0.5996	0.5933
		MRR	0.1680	0.1690	0.4834	0.4696	0.4041	0.4093	0.4995	0.4937
		Prec	0.0412	0.0397	0.0890	0.0881	0.0768	0.0776	0.0908	0.0908
		F1	0.0749	0.0722	0.1618	0.1602	0.1396	0.1411	0.1651	0.1651
	Cloth	HR	0.4147	0.3617	0.7241	0.7225	0.6794	0.6904	0.7702	0.7318
		NDCG	0.2395	0.2529	0.4178	0.4187	0.3986	0.3763	0.4577	0.4354
		MRR	0.1861	0.2199	0.3332	0.3250	0.3127	0.3391	0.3612	0.3443
		Prec	0.0415	0.0362	0.0724	0.0723	0.0679	0.0690	0.0770	0.0732
		F1	0.0754	0.0658	0.1316	0.1314	0.1235	0.1255	0.1400	0.1331
Task4	Elec	HR	0.5303	0.4333	0.7721	0.7781	0.6691	0.6913	0.7934	0.7889
		NDCG	0.3396	0.2907	0.4902	0.4912	0.3835	0.4258	0.5134	0.4920
		MRR	0.2809	0.2464	0.3936	0.3921	0.3549	0.3776	0.4268	0.4001
		Prec	0.0530	0.0433	0.0772	0.0778	0.0669	0.0691	0.0793	0.0789
		F1	0.0964	0.0787	0.1404	0.1415	0.1216	0.1256	0.1442	0.1435
	Cloth	HR	0.3925	0.3264	0.7698	0.7702	0.6299	0.6884	0.8261	0.8206
		NDCG	0.2266	0.2377	0.4552	0.4553	0.3259	0.3682	0.4950	0.4872
		MRR	0.1759	0.2109	0.3582	0.3583	0.2928	0.3192	0.3927	0.3840
		Prec	0.0393	0.0326	0.0770	0.0770	0.0630	0.0688	0.0826	0.0821
		F1	0.0714	0.0593	0.1400	0.1400	0.1145	0.1251	0.1502	0.1493
Task5	Book	HR	0.6515	0.5794	0.7128	0.6992	0.6519	0.6819	0.7323	0.7186
		NDCG	0.4057	0.3328	0.3801	0.4384	0.3387	0.4206	0.4615	0.4482
		MRR	0.3282	0.2572	0.3069	0.3526	0.2715	0.3391	0.3992	0.3640
		Prec	0.0652	0.0579	0.0713	0.0699	0.0652	0.0682	0.0732	0.0719
		F1	0.1185	0.1053	0.1296	0.1271	0.1185	0.1240	0.1331	0.1307
	Movie	HR	0.7488	0.6754	0.7443	0.7553	0.6660	0.6524	0.7773	0.7613
		NDCG	0.4762	0.4154	0.4765	0.4819	0.4047	0.3852	0.4996	0.4825
		MRR	0.3844	0.3351	0.3898	0.3894	0.3428	0.3015	0.4132	0.3955
		Prec	0.0749	0.0675	0.0744	0.0755	0.0666	0.0652	0.0777	0.0761
		F1	0.1362	0.1227	0.1353	0.1373	0.1211	0.1186	0.1413	0.1384
Task6	Movie	HR	0.6491	0.6548	0.7053	0.7019	0.6636	0.6798	0.7342	0.7168
		NDCG	0.3603	0.3952	0.4260	0.4175	0.3846	0.4039	0.4509	0.4349
		MRR	0.2713	0.3152	0.3403	0.3301	0.2881	0.3057	0.3532	0.3478
		Prec	0.0649	0.0655	0.0705	0.0702	0.0664	0.0680	0.0734	0.0717
		F1	0.1180	0.1191	0.1282	0.1276	0.1207	0.1236	0.1335	0.1304
	Music	HR	0.6937	0.6896	0.7691	0.7876	0.6673	0.6844	0.8029	0.7951
		NDCG	0.5064	0.4503	0.5508	0.5598	0.4657	0.5071	0.5843	0.5644
		MRR	0.4463	0.3751	0.4849	0.4879	0.4019	0.4507	0.5136	0.4919
		Prec	0.0694	0.0690	0.0769	0.0788	0.0667	0.0684	0.0803	0.0795
		F1	0.1262	0.1254	0.1398	0.1433	0.1213	0.1244	0.1460	0.1445

of K = [1, 3, 5, 7, 10], as illustrated in Fig. 7. As the length of the recommendation list (K) increases, model performance exhibits an upward trend. This is intuitive because a larger K implies that more items are recommended to the user, making the task simpler and more conducive to improved performance.

4.5.4. Impact of various information fusion methods

We employ three different aggregation methods, namely, element-wise sum, concatenation, and element-wise averaging, to combine obfuscated domain-common and domain-specific embeddings into the final user preferences. As illustrated in Fig. 8, concatenation yields the highest performance. This superiority can be attributed to the fact that

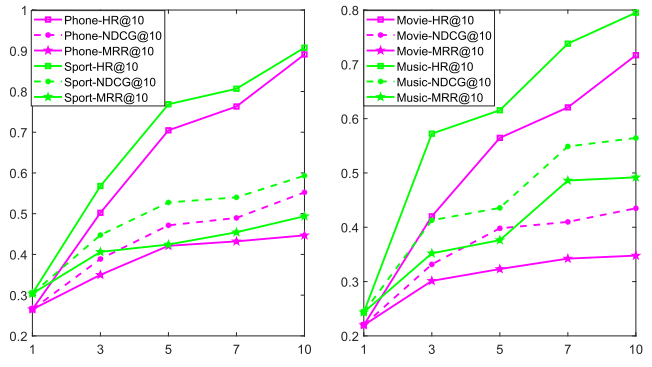


Fig. 7. Performance of different recommendation list length K.

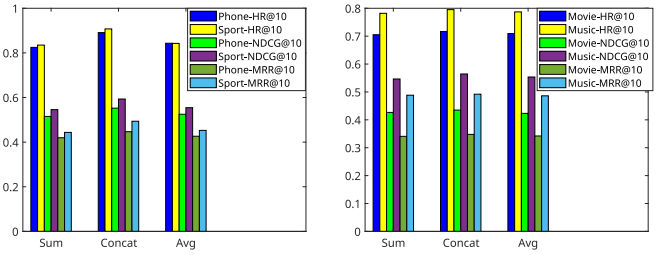


Fig. 8. Performance of different information fusion methods.

concatenation preserves all information and learns the complex relationships between domain-common and domain-specific embeddings more effectively.

4.6. Privacy vs performance (RQ5)

P2M2-CDR protects user privacy by utilizing LDP technology, where the parameter λ controls the strength of Laplace noise and the privacy budget. The degree to which noise is added significantly affects the privacy protection ability [35,38]. Adding more noise reduces data availability and the privacy budget, thereby enhancing privacy protection capabilities. However, excessive noise degrades the model's performance. Therefore, it is necessary to determine an optimal value of λ that balances the recommendation performance with privacy protection. To achieve this, we test P2M2-CDR's performance with different λ values [0.005, 0.01, 0.02, 0.03, 0.04], as shown in Fig. 9. We observe that a larger λ introduces more noise, enhancing user privacy protection but degrading model performance. The performance decreases significantly when λ is greater than 0.02. Therefore, we set λ to 0.01 to strike a balance between privacy preservation and maintaining acceptable performance.

4.7. Advantages and disadvantages

In this section, we aim to discuss the advantages and disadvantages of the proposed method. In this study, we propose a privacy-preserving framework for CDR that utilizes multi-modal data to decouple more informative embeddings and leverages LDP technology to protect user privacy during the knowledge transfer process. This method addresses the challenges faced by traditional CDR methods and achieves superior performance compared with other baselines. However, the proposed method has some limitations. First, the use of LDP technology may

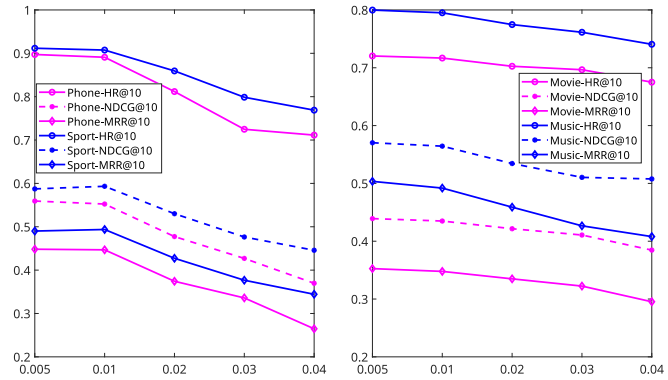


Fig. 9. Performance of different standard deviation λ.

degrade the model performance, necessitating a balance between privacy protection and recommendation accuracy. Second, our method relies on fully overlapping users to transfer knowledge across domains, which may lead to performance degradation when overlapping users are sparse.

5. Conclusion and future work

In this study, we propose a privacy-preserving framework with multi-modal data for CDR (P2M2-CDR). It contains a multi-modal disentangled encoder and a privacy-preserving decoder. The multi-modal encoder integrates multi-modal features to learn comprehensive user and item representations and disentangles user representations into more informative domain-common and domain-specific embeddings. The privacy-preserving decoder aims to introduce LDP to protect user privacy when transferring knowledge across domains. The experimental results for six CDR tasks demonstrate the effectiveness of our proposed model.

In the future, we plan to explore the following areas. First, we intend to investigate other prevalent privacy-preserving techniques, such as prototype-based FL, to enhance model performance while simultaneously protecting user privacy in CDR methods. Second, we will explore how to effectively utilize non-overlapping user information to solve the data-sparsity problem. Finally, We aim to develop more effective disentanglement techniques that can further refine the separation of domain-common and domain-specific embeddings, leading to more precise recommendations.

CRediT authorship contribution statement

Li Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Lei Sang:** Writing – review & editing. **Quangui Zhang:** Writing – review & editing. **Qiang Wu:** Writing – review & editing. **Min Xu:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the Australian Research Council, Australia (LP210100129).

References

[1] F. Zhu, Y. Wang, C. Chen, J. Zhou, L. Li, G. Liu, Cross-domain recommendation: challenges, progress, and prospects, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021), 2021, pp. 4721–4728, http://dx.doi.org/10.24963/ijcai.2021/639.

[2] G. Hu, Y. Zhang, Q. Yang, Conet: Collaborative cross networks for cross-domain recommendation, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 667–676.

[3] X. Xin, Z. Liu, C.-Y. Lin, H. Huang, X. Wei, P. Guo, Cross-domain collaborative filtering with review text, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[4] F. Zhu, Y. Wang, C. Chen, G. Liu, M. Orgun, J. Wu, A deep framework for cross-domain and cross-system recommendations, IJCAI '18, 2018, pp. 3711–3717.

[5] T. Man, H. Shen, X. Jin, X. Cheng, Cross-domain recommendation: An embedding and mapping approach, in: IJCAI, Vol. 17, 2017, pp. 2464–2470.

[6] C. Zhao, C. Li, R. Xiao, H. Deng, A. Sun, CATN: Cross-domain recommendation for cold-start users via aspect transfer network, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 229–238.

[7] P. Li, A. Tuzhilin, Ddtcdr: Deep dual transfer cross domain recommendation, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 331–339.

[8] F. Zhu, C. Chen, Y. Wang, G. Liu, X. Zheng, Dtcdr: A framework for dual-target cross-domain recommendation, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1533–1542.

[9] M. Liu, J. Li, G. Li, P. Pan, Cross domain recommendation via bi-directional transfer graph collaborative filtering networks, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 885–894.

[10] F. Zhu, Y. Wang, C. Chen, G. Liu, X. Zheng, A graphical and attentional framework for dual-target cross-domain recommendation, in: IJCAI, 2020, pp. 3001–3008.

[11] J. Cao, X. Lin, X. Cong, J. Ya, T. Liu, B. Wang, Disencdr: Learning disentangled representations for cross-domain recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 267–277.

[12] R. Zhang, T. Zang, Y. Zhu, C. Wang, K. Wang, J. Yu, Disentangled contrastive learning for cross-domain recommendation, in: International Conference on Database Systems for Advanced Applications, Springer, 2023, pp. 163–178.

[13] C. Chen, M. Zhang, C. Wang, W. Ma, M. Li, Y. Liu, S. Ma, An efficient adaptive transfer neural network for social-aware recommendation, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 225–234.

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

[15] J. McAuley, C. Targett, Q. Shi, A. Van Den Hengel, Image-based recommendations on styles and substitutes, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2015, pp. 43–52.

[16] Z. Cheng, Y. Ding, L. Zhu, M. Kankanhalli, Aspect-aware latent factor model: Rating prediction with ratings and reviews, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 639–648.

[17] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 639–648.

[18] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: Conference on Empirical Methods in Natural Language Processing, 2019.

[19] S. Berkovsky, T. Kuflik, F. Ricci, Cross-domain mediation in collaborative filtering, in: User Modeling 2007: 11th International Conference, UM 2007, Corfu, Greece, July 25-29, 2007. Proceedings 11, Springer, 2007, pp. 355–359.

[20] S. Tan, J. Bu, X. Qin, C. Chen, D. Cai, Cross domain recommendation based on multi-type media fusion, Neurocomputing 127 (2014) 124–134.

[21] J. Wang, J. Lv, Tag-informed collaborative topic modeling for cross domain recommendations, Knowl.-Based Syst. 203 (2020) 106119.

[22] A. Mnih, R.R. Salakhutdinov, Probabilistic matrix factorization, Adv. Neural Inf. Process. Syst. 20 (2007).

[23] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182.

[24] J. Ma, C. Zhou, P. Cui, H. Yang, W. Zhu, Learning disentangled representations for recommendation, Adv. Neural Inf. Process. Syst. 32 (2019).

[25] Y. Zheng, C. Gao, X. Li, X. He, Y. Li, D. Jin, Disentangling user interest and conformity for recommendation with causal embedding, in: Proceedings of the Web Conference 2021, 2021, pp. 2980–2991.

[26] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, T.-S. Chua, Disentangled graph collaborative filtering, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1001–1010.

[27] X. Guo, S. Li, N. Guo, J. Cao, X. Liu, Q. Ma, R. Gan, Y. Zhao, Disentangled representations learning for multi-target cross-domain recommendation, ACM Trans. Inf. Syst. 41 (4) (2023) 1–27.

[28] K. Menglin, J. Wang, Y. Pan, H. Zhang, M. Hou, C^2DR: Robust cross-domain recommendation based on causal disentanglement, in: Proceedings of the 17th ACM International Conference on Web Search and Data Mining, 2024, pp. 341–349.

[29] V. Perifanis, P.S. Efraimidis, Federated neural collaborative filtering, Knowl.-Based Syst. 242 (2022) 108441.

[30] Y. Wei, X. Fu, Q. Sun, H. Peng, J. Wu, J. Wang, X. Li, Heterogeneous graph neural network for privacy-preserving recommendation, in: 2022 IEEE International Conference on Data Mining, ICDM, IEEE, 2022, pp. 528–537.

[31] X. Yu, D. Zhan, L. Liu, H. Lv, L. Xu, J. Du, A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion, IEEE J. Biomed. Health Inf. 26 (5) (2021) 1928–1936.

[32] C. Gao, X. Chen, F. Feng, K. Zhao, X. He, Y. Li, D. Jin, Cross-domain recommendation without sharing user-relevant data, in: The World Wide Web Conference, 2019, pp. 491–502.

[33] D. Chai, L. Wang, K. Chen, Q. Yang, Secure federated matrix factorization, IEEE Intell. Syst. 36 (5) (2020) 11–20.

[34] H. Shin, S. Kim, J. Shin, X. Xiao, Privacy enhanced matrix factorization for recommendation with local differential privacy, IEEE Trans. Knowl. Data Eng. 30 (9) (2018) 1770–1782.

[35] C. Gao, C. Huang, Y. Yu, H. Wang, Y. Li, D. Jin, Privacy-preserving cross-domain location recommendation, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3 (1) (2019) 1–21.

[36] D. Yan, Y. Zhao, Z. Yang, Y. Jin, Y. Zhang, FedCDR: Privacy-preserving federated cross-domain recommendation, Digit. Commun. Netw. 8 (4) (2022) 552–560.

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

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

[39] X. Liao, W. Liu, X. Zheng, B. Yao, C. Chen, PPGenCDR: A stable and robust framework for privacy-preserving cross-domain recommendation, Proc. AAAI Conf. Artif. Intell. 37 (4) (2023) 4453–4461, http://dx.doi.org/10.1609/aaai.v37i4.25566.

[40] Y. Tan, G. Long, J. Ma, L. Liu, T. Zhou, J. Jiang, Federated learning from pre-trained models: A contrastive learning approach, Adv. Neural Inf. Process. Syst. 35 (2022) 19332–19344.

[41] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, C. Zhang, Fedproto: Federated prototype learning across heterogeneous clients, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 8432–8440.

[42] K. Zhou, H. Wang, W.X. Zhao, Y. Zhu, S. Wang, F. Zhang, Z. Wang, J.-R. Wen, S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1893–1902.

[43] F. Zhu, Y. Wang, J. Zhou, C. Chen, L. Li, G. Liu, A unified framework for cross-domain and cross-system recommendations, IEEE Trans. Knowl. Data Eng. (2021).

[44] J. Lu, G. Sun, X. Fang, J. Yang, W. He, A contrastive learning framework for dual-target cross-domain recommendation, in: Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 6332–6339.

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

[46] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).