Disentangled Contrastive Learning for Cross-Domain Recommendation

Ruohan Zhang, Tianzi Zang, Yanmin Zhu(✉), Chunyang Wang, Ke Wang, and Jiadi Yu

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{Zhangruohan,zangtianzi,yzhu,wangchy,onecall,jiadiyu}@sjtu.edu.cn

Abstract. Cross-Domain Recommendation (CDR) has been proved helpful in dealing with two bottlenecks in recommendation scenarios: data sparsity and cold start. Recent research reveals that identifying domain-invariant and domain-specific features behind interactions aids in generating comprehensive user and item representations. However, we argue that existing methods fail to separate domain-invariant and domain-specific representations from each other, which may contain noise and redundancy when treating domain-invariant representations as shared information across domains and harm recommendation performance. In this paper, we propose a novel **D**isentangled **C**ontrastive Learning for **C**ross-**D**omain **R**ecommendation framework (DCCDR) to disentangle domain-invariant and domain-specific representations to make them more informative. Specifically, we propose a *separate representation generation* component to generate separate domain-invariant and domain-specific representations for each domain. Next, We enrich the representations through multi-order collaborative information with GNNs. Moreover, we design a mutual-information-based contrastive learning objective to produce additional supervision signals for disentanglement and enhance the informativeness of disentangled representations by reducing noise and redundancy. Extensive experiments on two real-world datasets show that our proposed DCCDR model outperforms state-of-the-art single-domain and cross-domain recommendation approaches.

Keywords: Cross-domain Recommendation · Contrastive Learning · Disentangled Representation Learning · Graph Convolutional Networks

1 Introduction

Recommender systems have become key components for e-commerce and social media platforms, assisting users in accurately finding items they are potentially interested in amid overloaded information [36]. With the number of new items and users increasing, data sparsity and cold-start become two important issues hurting the efficacy of traditional recommender systems. Cross-domain recommendation (CDR) appears to be a solution to these two bottlenecks [35,42].

X. Wang et al. (Eds.): DASFAA 2023, LNCS 13944, pp. 163–178, 2023.
https://doi.org/10.1007/978-3-031-30672-3_11

Existing CDR approaches generally leverage rich information (i.e., ratings, feedback, tags, reviews) from source domains to improve recommendation performance in a target domain. Collective matrix factorization (CMF)-based methods [17,23,31] generate shared representations of users and items. Embedding and mapping-based methods [20] utilizing embedding mapping and dual knowledge transfer-based methods [9,14,15,40] have demonstrated efficacy in sharing knowledge across domains. Graph neural networks (GNNs)-based methods [4,12,32,37,41] in CDR utilize high-order collaborative information to improve performance. However, these methods consider that users have consistent interests across domains, so they only directly share common features when generating user representations, which ignores users may have different interests across domains. Recent approaches [3,10,16,33,38] consider domain-invariant and domain-specific features that are shared and distinct across domains [25]. They initialize two separate embeddings to represent domain-invariant and domain-specific features and transfer shared knowledge across domains to learn domain-invariant and domain-specific representations. Nonetheless, they fail to generate informative representations containing diverse semantics as it is difficult to distinguish between these two kinds of highly entangled features.

It is critical to separate domain-invariant and domain-specific representations from each other to enable them to contain more diverse semantic information. However, for existing CDR methods that identify domain-invariant and domain-specific features, their generated representations may have two main limitations as shown in Fig. 1, First, the domain-invariant representations may involve domain-specific features, which introduces noise caused by domain-specific features when treating domain-invariant representations as shared information to transfer, leading to the negative transfer problem [35]. Second, the domain-invariant and domain-specific representations may have some redundancy representing identical features, which decreases their expression ability greatly and results in sub-optimal results. Therefore, our method wants to separate domain-invariant and domain-specific representations to obtain more informative representations without noise and redundancy.

To reduce noise and redundancy contained in the generated domain-invariant and domain-specific representations and improve their informativeness, we are faced with the following challenges. The first challenge is *how to distinguish between domain-invariant and domain-specific features from user-item interactions*. Generally, user-item interactions are explicit ratings or implicit feedback in the datasets, so domain-invariant and domain-specific features behind user-item interactions are highly entangled, which makes it difficult to disentangle them. The second challenge is *how to enhance the informativeness of domain-invariant and domain-specific representations*. If domain-invariant and domain-specific representations are obtained, they may involve noise caused by irrelevant features, which hinders the effectiveness of sharing knowledge across domains and hurts recommendation performance. The generated representations may have some redundant information, which reduces the informativeness of representations.

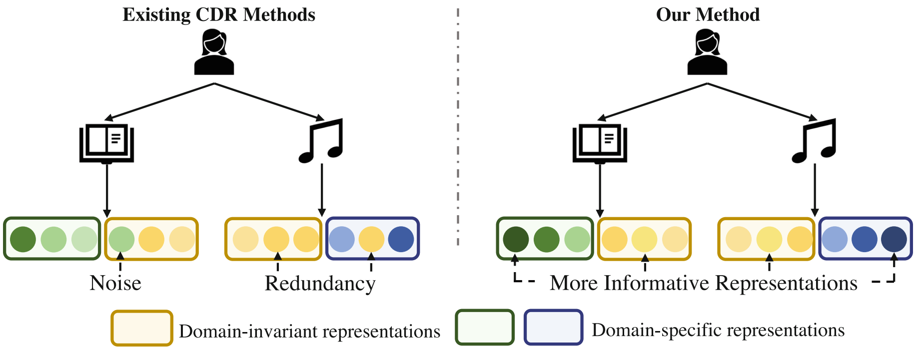


Fig. 1. Existing CDR models generate domain-invariant and domain-specific representations which contain noise and redundancy, whereas our methods generate more informative representations.

Therefore, it is difficult to remove noise and redundancy from the generated representations to make them more informative.

In this paper, we develop a new model called **D**isentangled **C**ontrastive Learning for **C**ross-**D**omain **R**ecommendation (DCCDR) to disentangle domain-invariant and domain-specific representations. To tackle the aforementioned challenges, we propose a *disentangled contrastive learning module* to generate more informative domain-invariant and domain-specific representations. To address the first challenge, we develop a *separate representation generation* component to generate separate domain-invariant and domain-specific representations. To enrich representations, we perform a GNNs-based approach to utilize high-order collaborative information in the user-item interaction graph. To deal with the second challenge, we develop a *representation informativeness enhancement* component to supervise the disentanglement and enhance the informativeness of representations by reducing noise and redundancy. Specifically, the contrastive learning objective maximizes the mutual information between domain-invariant representations of users across domains to reduce noise. To make disentangled representations contain more diverse semantics, the objective minimizes the mutual information between domain-invariant and domain-specific representations within domains and that between domain-specific representations across domains. Finally, we concatenate the domain-invariant and domain-specific representations to generate the final ones and predict the probability of given user-item pairs. In summary, the main contributions of this paper are as follows:

- We emphasize the importance of distinguishing domain-invariant and domain-specific features in cross-domain recommendation. A novel model DCCDR is proposed to disentangle domain-invariant and domain-specific representations to enable them to be more informative and contain diverse semantics.
- We develop the *representation informativeness enhancement* to supervise the disentanglement and enhance the informativeness of disentangled representations by reducing noise and redundancy. A mutual-information-based

contrastive learning objective is designed to add supervision signals for model training and representation enhancement.
- We conduct extensive experiments on real-world Amazon and Douban datasets. Comprehensive results demonstrate that our model significantly outperforms the state-of-the-art methods of cross-domain recommendation.

2 Related Work

2.1 Cross-Domain Recommendation

Cross-domain recommendation (CDR), which takes advantage of the abundant knowledge in source domains, often improves recommendation performance in a sparse target domain. Traditional approaches factorize rating matrices jointly and capture common user preferences [17,23]. As deep learning techniques gain popularity, methods use mapping functions [20], domain adaption skills [5,34], dual knowledge transfer mechanism [9,14,40], and graph neural networks (GNNs) [4,12,32,37,41] to transfer or leverage shared knowledge across domains. Recent approaches consider domain-invariant and domain-specific features when sharing knowledge across domains [3,10,16,33,38]. For instance, Zhao et al. propose MSDCR [38] to learn domain-specific and domain-invariant user preferences at the aspect level by transferring the user's complementary aspect preferences across domains. However, existing CDR approaches fail to separate domain-invariant and domain-specific representations from each other, which introduces noise and redundancy. Unlike prior research, Our DCCDR aims to disentangle domain-invariant and domain-specific representations to enhance their informativeness.

2.2 Disentangled Learning in Recommendation

Disentangled Learning [1] is proposed to learn distinct representations from multiple latent factors that influence the data, which is well aligned with recommendation tasks. Disentangled learning has proved effective in single-domain recommendation [18], sequential recommendation [19,39], and social recommendation [13] Researchers have shown that GNNs are effective to learn disentangled representations from graphs for recommendation. GNN-based disentangled methods leverage the interaction graph [26], the heterogeneous graph [27], and the knowledge graph [29]. Recently, learning disentangled representations has been introduced in CDR tasks [2]. Yet, the aforementioned works fail to separate domain-invariant and domain-specific representations from each other and enhance their informativeness, which leads to sub-optimal recommendation performance in CDR. These short-comings can be addressed by our proposed DCCDR model, which disentangles domain-invariant and domain-specific representations with generated supervised signals to make them contain more diverse semantics.

3 Notations and Problem Definition

Let D^A and D^B denote two distinct domains which share the same set of users denoted as $\mathcal{U}$. $\mathcal{I}^A$ and $\mathcal{I}^B$ denote non-overlapped sets of items in domain D^A

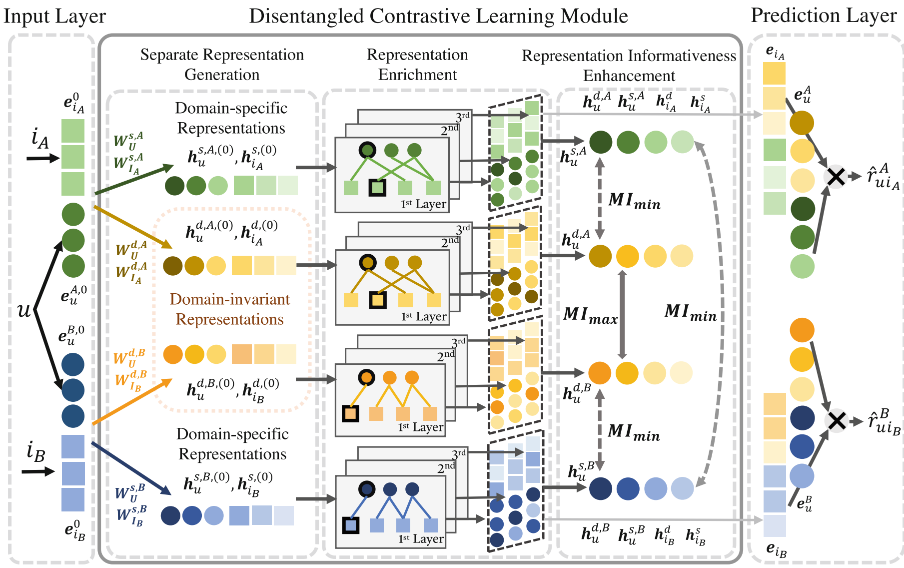


Fig. 2. An architecture overview of our model **DCCDR**. The core module of DCCDR is the *Disentangled Contrastive Learning Module*, which contains three key components: (1) the *Separate Representation Generation*, (2) the *Representation Enrichment*, and (3) the *Representation Informativeness Enhancement*.

and D^B, among which no item is in common. $|\mathcal{U}|$, $|\mathcal{I}^A|$, and $|\mathcal{I}^B|$ are the number of shared users and items in each domain. $\mathbf{G}^A = (\mathcal{U},\ \mathcal{I}^A,\ \mathbf{R}^A)$ and $\mathbf{G}^B = (\mathcal{U},\ \mathcal{I}^B,\ \mathbf{R}^B)$ denote the interaction graph in D^A and D^B separately, where $\mathbf{R}^A \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}^A|}$ and $\mathbf{R}^B \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}^B|}$ are user-item interaction matrices. $\mathbf{R}_{ui} = 1$ indicates an observed interaction between the user u and item i, otherwise 0.

With two domains D^A and D^B and a set of common users $\mathcal{U}$, two sets of non-overlapped items $\mathcal{I}^A$, $\mathcal{I}^B$, and corresponding interaction matrices $\mathbf{R}^A$ and $\mathbf{R}^B$ given, We consider Top-N recommendation with implicit feedback for all users in each domain. That is to say, we recommend a set of items $\mathcal{I}^a \subset \mathcal{I}^A$, $\mathcal{I}^b \subset \mathcal{I}^B$ that users have not interacted with but are most likely to be interested in to improve the recommendation performance in both domains simultaneously.

4 DCCDR

The main structure of our **D**isentangled **C**ontrastive learning networks for **C**ross-**D**omain **R**ecommendation (DCCDR) is shown in Fig. 2, which contains the *input layer*, the *disentangled contrastive learning module* and the *prediction layer*. In the following, we will introduce it in detail.

4.1 Input Layer

Considering that two domains share the same user set, the inputs of our proposed DCCDR model in the two domains are the user-item pairs $(u,\ i_A)$ and $(u,\ i_B)$. We denote both users and items by one-hot encodings, i.e., $\mathbf{x}_u \in \{0,1\}^{|\mathcal{U}|}$, $\mathbf{x}_{i_A} \in \{0,1\}^{|\mathcal{I}^{\mathcal{A}}|}$, $\mathbf{x}_{i_B} \in \{0,1\}^{|\mathcal{I}^{\mathcal{B}}|}$. We generate the initialized embeddings through embedding matrices $\mathbf{E}_{U^A}, \mathbf{E}_{U^B}, \mathbf{E}_{I^A}, \mathbf{E}_{I^B}$:

$$\mathbf{e}_u^{A,0} = \mathbf{E}_{U^A}^T \mathbf{x}_u, \quad \mathbf{e}_u^{B,0} = \mathbf{E}_{U^B}^T \mathbf{x}_u, \quad \mathbf{e}_{i_A}^0 = \mathbf{E}_{I^A}^T \mathbf{x}_{i_A}, \quad \mathbf{e}_{i_B}^0 = \mathbf{E}_{I^B}^T \mathbf{x}_{i_B}, \tag{1}$$

where $\mathbf{E}_{U^A} \in \mathbb{R}^{|\mathcal{U}|\times d}, \mathbf{E}_{U^B} \in \mathbb{R}^{|\mathcal{U}|\times d}, \mathbf{E}_{I^A} \in \mathbb{R}^{|\mathcal{I}^{\mathcal{A}}|\times d}, \mathbf{E}_{I^B} \in \mathbb{R}^{|\mathcal{I}^{\mathcal{B}}|\times d}$, $\mathbf{e}_u^0$ and $\mathbf{e}_i^0$ denote the generated initial embeddings of the user u and the item i respectively, and d is the dimension of all embeddings.

4.2 Disentangled Contrastive Learning Module

This module is the core of our proposed DCCDR model, which aims to disentangle domain-invariant and domain-specific representations to make them more separate and informative. It is crucial to disentangle to reduce noise caused by irrelevant features when sharing domain-invariant representations across domains. It will also help to reduce the redundancy of these two representations to obtain informative representations with more diverse semantics. There are three problems faced with the disentanglement. The first one is how to extract domain-invariant and domain-specific features and generate two separate representations. The second one is how to leverage collaborative information to enrich representations. The third one is how to enhance the informativeness of disentangled representations. We propose a *separate representation generation* component, a *representation enrichment* component and a *representation informativeness enhancement* component to tackle these problems.

Separate Representation Generation. We employ latent space projection to generate separate domain-invariant and domain-specific representations. Specifically, for the user u, different from other recommendation methods [9,23,40] which capture a uniform user interest and generate a holistic representation, we extract domain-invariant and domain-specific features as two independent parts of user interest. Formally, we project the original user embedding $\mathbf{e}_u^0$ into different latent spaces:

$$\mathbf{h}_u^{d,(0)} = \sigma(\mathbf{W}_U^d \mathbf{e}_u^0), \qquad \mathbf{h}_u^{s,(0)} = \sigma(\mathbf{W}_U^s \mathbf{e}_u^0), \tag{2}$$

where d and s denote the domain-invariant and domain-specific latent space respectively, W_U is the projection matrix and $\sigma(\cdot)$ is the activation function. The representations of each u would be composed of two parts, i.e., $\mathbf{e}_u^{(0)} = [\mathbf{h}_u^{d,(0)},\ \mathbf{h}_u^{s,(0)}]$, where $\mathbf{h}_u^{d,(0)}$ and $\mathbf{h}_u^{s,(0)}$ denote the domain-invariant and domain-specific representation of the user u. Analogously, we can generate separate representations $\mathbf{h}_i^{d,(0)},\ \mathbf{h}_i^{s,(0)}$ for item i.

Representation Enrichment. To enrich the representations, we leverage high-order collaborative information in the interaction graphs due to the effectiveness of GNNs [6,11,24] proved in recent studies. Research has shown that applying the embedding propagation mechanism on graph structure can extract useful information by aggregating information from neighbors and updating original nodes. The basic idea of Graph Convolutional Networks (GCNs) [11] is to learn node representations by smoothing features over the graph. It performs the following neighborhood aggregation iteratively to achieve a new representation of a target user node u with K convolutional layers:

$$e_u^{(k+1)} = AGG(e_u^{(k)}, e_i^{(k)} : i \in \mathcal{N}_u), \tag{3}$$

where k indicates the current convolutional layer, $e_u^{(k)}$ denotes the user embedding in the k^{th} layer, $\mathcal{N}_u$ is the set of neighbors of u in the interaction graph and AGG symbolizes the chosen aggregation strategy.

We adopt LightGCN [7] to enrich domain-invariant and domain-specific representations in each domain due to its low number of parameters. Take the domain-invariant representations in domain D^A as an example and the aggregation for a user u and an item i_A can be summarized as follows:

$$\mathbf{h}_u^{d,A,(k+1)} = \sum_{i \in \mathcal{N}_u^A} \frac{1}{\sqrt{|\mathcal{N}_u^A||\mathcal{N}_i|}} \mathbf{h}_i^{d,(k)}, \quad \mathbf{h}_{i_A}^{d,(k+1)} = \sum_{u \in \mathcal{N}_{i_A}} \frac{1}{\sqrt{|\mathcal{N}_u^A||\mathcal{N}_{i_A}|}} \mathbf{h}_u^{d,A,(k)}. \tag{4}$$

To achieve more comprehensive representations from multi-order neighbors, we combine the embeddings learned in each layer since different embedding layers capture different semantics as follows:

$$\mathbf{h}_u^{d,A} = \mathbf{h}_u^{d,A,(0)} \mid\mid ... \mid\mid \mathbf{h}_u^{d,A,(K)}, \quad \mathbf{h}_{i_A}^{d} = \mathbf{h}_{i_A}^{d,(0)} \mid\mid ... \mid\mid \mathbf{h}_{i_A}^{d,(K)}, \tag{5}$$

where $||$ denotes the concatenation operation. The enrichment of domain-specific representations can be performed similarly. This enrichment process is the same for both D^A and D^B.

Representation Informativeness Enhancement. To supervise the disentanglement and enhance the informativeness of disentangled representations, we utilize self-supervised learning (SSL) [22,28,30] to generate extra supervision signals for model training. Specifically, this component adopts the mutual information (MI) maximization mechanism [21] to construct a contrastive learning objective. The core idea of mutual information maximization is to select *positive pairs* whose MI should be maximized (i.e., MI_{max} in Fig. 2) and *negative pairs* whose MI should be minimized (i.e., MI_{min} in Fig. 2).

To reduce noise caused by irrelevant features when treating domain-invariant user representations as knowledge to be shared across domains, we treat the two domain-invariant user representations of the same user as a *positive pair* and minimize the difference between them because intuitively users' domain-invariant features are shared and consistent. To reduce redundancy of the disentangled representations and enhance their informativeness, we propose two negative pairs for our designed contrastive learning objective. First, if domain-invariant and domain-specific representations contain distinguishable semantics in each domain, all the shared knowledge will be squeezed into domain-invariant representations. Hence, the domain-specific and domain-invariant representations of the same user in each domain are treated as a *negative pair*. Second, based on the assumption that all the common knowledge is supposed to be represented in domain-invariant representations, it is convincing that the domain-specific representations of users across domains should have little mutual information and can be considered as another *negative pair*. Therefore, the contrastive learning objective is determined as follows:

$$\mathcal{L}_{CL} = -log\frac{exp(f(\mathbf{h}_u^{d,A},\ \mathbf{h}_u^{d,B}))/\tau)}{\begin{array}{c}exp(f(\mathbf{h}_u^{d,A},\ \mathbf{h}_u^{d,B}))/\tau) + exp(f(\mathbf{h}_u^{d,A},\ \mathbf{h}_u^{s,A}))/\tau) + \\ exp(f(\mathbf{h}_u^{d,B},\ \mathbf{h}_u^{s,B}))/\tau) + exp(f(\mathbf{h}_u^{s,A},\ \mathbf{h}_u^{s,B}))/\tau)\end{array}}, \tag{6}$$

where $f(\cdot)$ is a function measuring the mutual information contained in representation and τ is a hyper parameter for softmax temperature. Here we utilize the cosine similarity and other functions can be adopted.

$$f(\mathbf{h}_u^d,\ \mathbf{h}_u^s) = cos(\mathbf{h}_u^d,\ \mathbf{h}_u^s) = \frac{{\mathbf{h}_u^d}^T \mathbf{h}_u^s}{||\mathbf{h}_u^d||\ ||\mathbf{h}_u^s||}. \tag{7}$$

4.3 Prediction Layer

Through the *disentangled contrastive learning module*, we can obtain disentangled domain-invariant and domain-specific representations for users and items. To make representations more comprehensive, we adopt a concatenation operation to fuse the disentangled representations as follows:

$$\mathbf{e}_u^A = \mathbf{h}_u^{d,A}\ ||\ \mathbf{h}_u^{s,A}, \quad \mathbf{e}_{i_A} = \mathbf{h}_{i_A}^d\ ||\ \mathbf{h}_{i_A}^s, \tag{8}$$

where $\mathbf{e}_u^A$ and $\mathbf{e}_{i_A}$ are the final representation for user u and item i_A in D^A respectively. Finally, we utilize the dot product to calculate the probability of the interaction between u and i_A in D^A:

$$\hat{r}_{ui_A}^A = \hat{y}^A(u,\ i_A) = \sigma({\mathbf{e}_u^A}^T \mathbf{e}_{i_A}), \tag{9}$$

where σ is the sigmoid function to map real multiplication results to probability of interactions. Note that $\hat{r}_{ui_B}^B$ can be achieved through a similar process.

4.4 Model Training

We employ the *Bayesian Personalized Ranking* (BPR) loss, which is a typical pairwise loss encouraging a higher predicted probability of an observed interaction compared with unobserved ones.

$$\mathcal{L}_{BPR}^{'} = -\sum_{(u,i)\in\mathbf{R}^+,(u,j)\in\mathbf{R}^-} ln\ \sigma(\hat{r}_{ui} - \hat{r}_{uj}), \tag{10}$$

where $'$ denotes the chosen domain D^A or D^B to estimate corresponding BPR loss, i.e., $\mathcal{L}^A_{BPR}$ and $\mathcal{L}^B_{BPR}$. $\mathbf{R}^+$ is the set of observed interactions between users and items while $\mathbf{R}^-$ is the sampled set of unobserved interactions. Considering the determined disentangled contrastive learning objective in Formula (6), the total joint loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}^A_{BPR} + \mathcal{L}^B_{BPR} + \beta\mathcal{L}_{CL} + \lambda||\Theta||^2_2, \tag{11}$$

where β is the weight of $\mathcal{L}_{CL}$ and λ controls the L_2 regularization on the parameter set Θ to prevent overfitting.

5 Experiments

In this section, we discuss the experimental setup. We adopt two real-world datasets to conduct experiments and demonstrate the effectiveness of the proposed model. We expect to find answers to the following research questions.

- RQ1: Can our proposed model outperform other state-of-the-art approaches?
- RQ2: Do our designs aid in enhancing the performance of our model?
- RQ3: How do various hyper-parameter values affect our performance?
- RQ4: Are the representations we learned really disentangled?

5.1 Experimental Settings

Datasets We use two real-world datasets in our experiments to assess the performance of our proposed approach. The first dataset, **Amazon**[1], is the most often utilized for CDR. We select four domains for evaluation: "Movies and TV," "Digital Music," "Cell Phones and Accessories," and "Electronics" (abbreviated "Amazon-Movie," "Amazon-Music," "Amazon-Cell," and "Amazon-Elec"). The second dataset is the **Douban** dataset, which is crawled from the Douban website[2], a prominent online social network. There are three domains denoted as "Douban-Movie", "Douban-Music" and "Douban-Book", respectively. For each dataset, we treat the ratings of 4–5 as positive samples and others as negative ones, where each interaction is marked as 1 otherwise 0. For each task, we select the common users across both domains who have more than 3 interactions in each domain and limit each domain to less than or equal to 10000 items. Table 1 summarizes the detailed statistics of the datasets.

Evaluation Metrics. We adopt the leave-one-out strategy to evaluate our approach and baselines [40]. We randomly select one interaction as the test item for each user and determine hyper parameters by randomly sampling another interaction as the validated item. We randomly choose 999 negative items that are not interacted with by the user and rank the test item among the combined

[1] http://jmcauley.ucsd.edu/data/amazon/.
[2] https://www.douban.com.

Table 1. Experimental datasets and tasks.

Dataset	Domain	#Users	#Items	#Interaction	sparsity
Amazon	Movie	6995	10000	215299	99.69%
	Music	6995	10000	162779	99.77%
	Cell	7988	8455	86784	99.87%
	Elec	7988	9513	51398	99.93%
Douban	Movie	4494	10000	2038134	95.46%
	Book	4494	10000	303329	99.33%
	Music	6529	10000	704858	98.92%
	Book	6529	10000	467335	99.28%

1000 items because we are interested in the top-N recommendation tasks. This process is repeated five times, and the average ranking results are displayed. The recommendation performance is evaluated by two metrics: *Hit Ratio* (HR) and *Normalized Discounted Cumulative Gain* (NDCG). HR examines whether the test item is in the top-N ranking list, and NDCG measures the ranking quality by assigning higher scores to hits at top ranks [16].

Comparison Methods. We compare our proposed DCCDR model with both single-domain and cross-domain recommendation methods.

- **Single-domain recommendation. NeuMF** [8] combines matrix factorization (MF) and Deep Neural Networks (DNNs) to model user-item latent interactions. **LightGCN** [7] proposes a simplified neighborhood aggregation through normalized sum to generate representations.
- **Cross-domain recommendation. CMF** [23] jointly factorizes matrices and shares the latent factors of overlapped users. **DeepAPF** [33] captures both cross-site common and site-specific interests with weights learned by the attentional network. **CoNet** [9] introduces cross-connection units to conduct a dual knowledge transfer. **PPGN** [37] constructs a cross-domain preference matrix to maintain the cross-domain interactions and captures high-order connections. **BiTGCF** [16] conducts a bi-direction transfer learning through graph collaborative filtering. **MSDCR** [38] enhances domain-specific aspect preferences through adversarial training to form comprehensive preferences.

Experiment Setup. We use PyTorch to develop DCCDR[3], and all experiments are run on an NVIDIA TITAN Xp GPU. The dimension d of both the user and item embeddings is set to 64. The number of graph convolutional layers is set to 3. We set the disentangled contrastive learning objective's weight β to 0.001. With a learning rate of 0.001, we use the Adam optimizer in a mini-batch mode to update parameters. The batch size is set to 1024. Furthermore, to prevent

[3] https://github.com/wangshanyw/DCCDR.

Table 2. Performance comparison between **DCCDR** and different methods. Best baselines are underlined. $\star$ indicates the statistical significance for $p \leq 0.01$ compared with the best baseline method based on the paired t-test.

Metric	H@2	H@5	N@5	H@2	H@5	N@5	H@2	H@5	N@5	H@2	H@5	N@5
Datasets	Amazon-Movie			Amazon-Music			Amazon-Cell			Amazon-Elec		
NeuMF	0.045	0.090	0.057	0.064	0.117	0.076	0.051	0.080	0.056	0.047	0.079	0.055
LightGCN	0.120	0.232	0.141	0.150	0.229	0.171	0.141	0.260	0.170	0.155	0.247	0.180
CMF	0.085	0.150	0.101	0.106	0.176	0.124	0.127	0.192	0.141	0.109	0.159	0.119
DeepAPF	0.067	0.119	0.080	0.095	0.151	0.106	0.072	0.112	0.080	0.058	0.095	0.067
CoNet	0.153	0.237	0.170	0.151	0.235	0.169	0.119	0.176	0.132	0.122	0.179	0.135
PPGN	0.178	0.268	0.200	0.241	0.356	0.268	0.307	0.436	0.336	0.202	0.306	0.226
BiTGCF	0.179	0.283	0.204	0.230	0.351	0.257	0.250	0.359	0.272	0.233	0.338	0.258
MSDCR	0.215	0.328	0.245	0.298	0.420	0.311	0.410	0.548	0.422	0.282	0.415	0.310
DCCDR	$0.259^\star$	$0.376^\star$	$0.285^\star$	$0.330^\star$	$0.471^\star$	$0.359^\star$	$0.476^\star$	$0.616^\star$	$0.493^\star$	$0.315^\star$	$0.465^\star$	$0.347^\star$
Improv.	**21%**	**15%**	**16%**	**11%**	**12%**	**16%**	**16%**	**12%**	**17%**	**12%**	**12%**	**12%**
Datasets	Douban-Movie			Douban-Book			Douban-Music			Douban-Book		
NeuMF	0.088	0.160	0.105	0.092	0.173	0.112	0.096	0.175	0.115	0.097	0.178	0.118
LightGCN	0.093	0.166	0.112	0.135	0.227	0.155	0.122	0.119	0.143	0.119	0.160	0.158
CMF	0.060	0.106	0.070	0.086	0.152	0.101	0.079	0.142	0.096	0.086	0.152	0.103
DeepAPF	0.086	0.160	0.107	0.132	0.225	0.154	0.110	0.197	0.132	0.119	0.203	0.140
CoNet	0.237	0.357	0.265	0.243	0.357	0.264	0.134	0.228	0.170	0.151	0.221	0.186
PPGN	0.097	0.176	0.116	0.149	0.244	0.168	0.122	0.210	0.146	0.128	0.218	0.150
BiTGCF	0.084	0.160	0.105	0.184	0.277	0.203	0.146	0.224	0.167	0.154	0.227	0.176
MSDCR	0.260	0.360	0.285	0.250	0.354	0.275	0.180	0.217	0.190	0.148	0.201	0.163
DCCDR	$0.309^\star$	$0.386^\star$	$0.314^\star$	$0.288^\star$	$0.375^\star$	$0.307^\star$	$0.198^\star$	$0.241^\star$	$0.211^\star$	$0.175^\star$	$0.238^\star$	$0.208^\star$
Improv.	**19%**	**7%**	**10%**	**15%**	**6%**	**12%**	**10%**	**6%**	**11%**	**13%**	**5%**	**12%**

overfitting in graph convolution, we implement the message dropout mechanism during propagation with a dropout of 0.4 in training and disable it during testing.

5.2 Performance Comparison (RQ1)

Table 2 summarizes the results of our experiments on four tasks using HR@2 (H@2), HR@5 (H@5), and NDCG@5 (N@5). We can see that CDR methods (such as CMF, CoNet, and MSDCR) outperform single-domain recommendation approaches in general (i.e. NeuMF). This demonstrates how useful data across domains help improve recommendation performance. Moreover, in most circumstances, GNN-based recommender systems surpass non-graph recommendation methods (e.g., LightGCN vs NeuMF, PPGN, vs CoNet, etc.). This demonstrates the efficacy of using the graph to model high-order relationships. Additionally, among previous methods identifying domain-invariant and domain-specific features, MSDCR outperforms other approaches (i.e. DeepAPF, BiTGCF) but is still weaker than our model, which disentangles domain-invariant and domain-specific representations from each other to make them more informative. Our proposed DCCDR model performs optimally on all tasks. Over four pairs of

Table 3. Results of ablation study.

Metric	HR@2	HR@5	NDCG@5	HR@2	HR@5	NDCG@5
Variants	Amazon-Movie			Amazon-Music		
w/o.g	0.1459	0.2011	0.1620	0.1495	0.2106	0.1704
w/o.cl	0.1609	0.2213	0.1838	0.1740	0.2353	0.2165
w/o.dt	0.2357	0.3410	0.2630	0.3150	0.4647	0.3471
DCCDR	**0.2588**	**0.3763**	**0.2854**	**0.3297**	**0.4712**	**0.3590**

tasks, the average performance improvement is 15.2%, 13.5%, 11.5%, and 9.5%, demonstrating the effectiveness of our design.

5.3 Ablation Study (RQ2)

We conduct an ablation study to compare DCCDR with three variants to evaluate the effectiveness of each designed module in DCCDR. *w/o.srg* replaces the *separate representation generation* component with two initialized embeddings for domain-invariant and domain-specific representations for each domain. *w/o.g* replaces LightGCN in the *representation enrichment* component with matrix factorization (MF). *w/o.cl* removes the contrastive learning objective from the joint loss (i.e. $\beta = 0$). The outcomes of the experiments are shown in Table 3.

We can see that without GNNs leveraging multi-hop connections in graphs, *w/o.g* performs the poorest, with an average drop of 49.32%. This demonstrates the importance of modeling high-order relationships to enrich user and item representations. The recommendation performance of *w/o.cl* decreases by 41.93% on average, which demonstrates contrastive learning helps share information across domains and learn more informative disentangled representations. The drop of *w/o.dt* (i.e. 1.38–9.38%) demonstrates the importance of separating domain-invariant and domain-specific representations. Our DCCDR greatly outperforms all variants in terms of HR@2, HR@5, and NDCG@5, indicating each designed component truly contributes to performance improvement.

5.4 Impact of Hyper-parameter Settings (RQ3)

Impact of the weight β **of** $\mathcal{L}_{CL}$**.** We first investigate the impact of the weight β of contrastive learning objective $\mathcal{L}_{CL}$. Taking "Amazon-Cell↔Amazon-Elec" and "Douban-Music↔Douban-Book" as examples, we select {0.0001, 0.0005, 0.001, 0.002, 0.005, 0.01} as β respectively, and show experimental results in Fig 3. The recommendation performance improves as β increases and peaks at 0.001. Then the recommendation performance decreases as β becomes larger. We think this is because when β reaches 0.001, it keeps a good balance between

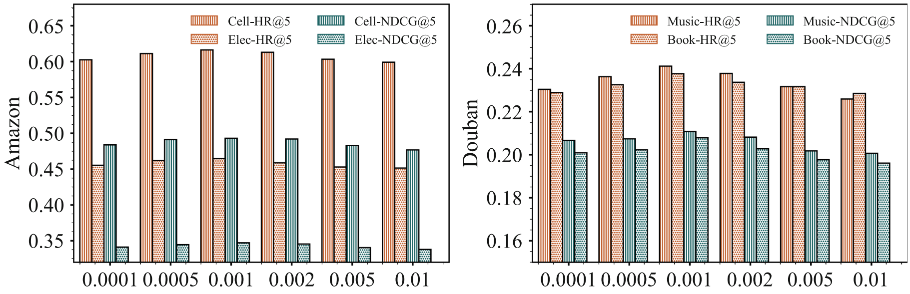


Fig. 3. Impact of the weight β of $\mathcal{L}_{CL}$.

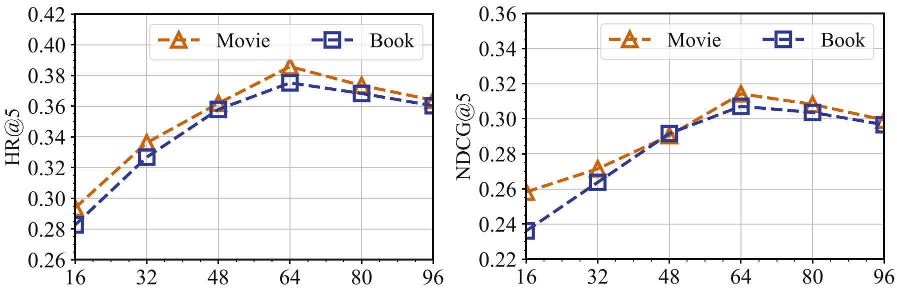


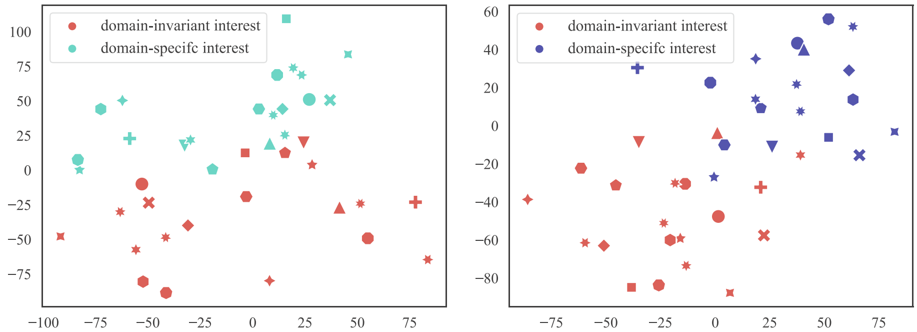
Fig. 4. Impact of the embedding size d.

BPR loss and contrastive learning objective, which helps learn more informative representations. Therefore, we set β as 0.001.

Impact of the Embedding Size *d*. We then investigate the impact of the dimension of user/item representations. Taking the "Douban-Movie↔ Douban-Music" task as an example, we range the embedding size within {16, 32, 48, 64, 80, 96} and plot the results in Fig. 4. In the beginning, as the size of the embeddings increases, both HR@5 and NDCG@5 increase. This may be because a relatively large size of embeddings helps represent more semantics. The recommendation performance peaks when the dimension reaches 64 and declines when the size of embeddings is larger than 64. This may be caused by the overfitting of the model. Therefore, in our experiments, we set the embedding size to 64.

5.5 Visualization (RQ4)

We visualize the disentangled user representations to see if domain-invariant and domain-specific representations are separate from each other and contain diverse semantics. In the "Douban-Music↔Douban-Book" task, we randomly select 20 groups of disentangled user representations in both domains $\mathbf{h}_u^{d,A}$, $\mathbf{h}_u^{s,A}$, $\mathbf{h}_u^{d,B}$, $\mathbf{h}_u^{s,B}$. Using the t-SNE technique, we project the high-dimensional representations into 2D space. The visualization results are displayed in Fig. 5, where red



(a) Douban-Music (b) Douban-Book

Fig. 5. Visualization of disentangled user representations.

represents domain-invariant interest and blue represents domain-specific one. Distinct shapes (i.e., squares and circles) represent different groups of disentangled user representations. We can see that the clustering centers of domain-invariant and domain-specific interests are separate, suggesting that interactions have distinguishable domain-invariant and domain-specific properties. Moreover, the representations of the same user (i.e., the same shape) are far apart, indicating that our model can disentangle user representations.

6 Conclusion and Future Work

In this paper, we propose DCCDR to disentangle domain-invariant and domain-specific representations to make them more informative. A *separate representation generation* component is designed to generate separate domain-invariant and domain-specific representations. We design a mutual-information based contrastive learning objective to generate supervised signals and enhance representation informativeness by reducing noise and redundancy. Extensive experiments demonstrate the effectiveness of our proposed model. Currently, we only disentangle representations based on the interaction graph and the disentanglement can be applied to other graphs (e.g. social graph). In the future, we will research more efficient ways to generate disentangled representations based on various types of information and further improve recommendation performance.

Acknowledgement. This research is supported in part by National Science Foundation of China (No. 62072304, No. 61772341, No. 61832013, No. 62172277, No. 62272320), Shanghai Municipal Science and Technology Commission (No. 19510760500, No. 21511104700, No. 19511120300), and Zhejiang Aoxin Co. Ltd.

References

1. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
2. Cao, J., Lin, X., Cong, X., Ya, J., Liu, T., Wang, B.: DisenCDR: learning disentangled representations for cross-domain recommendation. In: SIGIR, pp. 267–277 (2022)
3. Chen, C., et al.: An efficient adaptive transfer neural network for social-aware recommendation. In: SIGIR, pp. 225–234 (2019)
4. Cui, Q., Wei, T., Zhang, Y., Zhang, Q.: HeroGRAPH: a heterogeneous graph framework for multi-target cross-domain recommendation. In: ORSUM@ RecSys (2020)
5. Duan, L., Tsang, I.W., Xu, D., Chua, T.S.: Domain adaptation from multiple sources via auxiliary classifiers. In: ICML, pp. 289–296 (2009)
6. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
7. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: simplifying and powering graph convolution network for recommendation. In: SIGIR, pp. 639–648 (2020)
8. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.: Neural collaborative filtering. In: WWW, pp. 173–182 (2017)
9. Hu, G., Zhang, Y., Yang, Q.: CoNet: collaborative cross networks for cross-domain recommendation. In: CIKM, pp. 667–676 (2018)
10. Huang, L., Zhao, Z.L., Wang, C.D., Huang, D., Chao, H.Y.: LSCD: low-rank and sparse cross-domain recommendation. Neurocomputing **366**, 86–96 (2019)
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
12. Li, J., Peng, Z., Wang, S., Xu, X., Yu, P.S., Hao, Z.: Heterogeneous graph embedding for cross-domain recommendation through adversarial learning. In: Nah, Y., Cui, B., Lee, S.-W., Yu, J.X., Moon, Y.-S., Whang, S.E. (eds.) DASFAA 2020. LNCS, vol. 12114, pp. 507–522. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59419-0_31
13. Li, N., Gao, C., Jin, D., Liao, Q.: Disentangled modeling of social homophily and influence for social recommendation. In: TKDE (2022)
14. Li, P., Tuzhilin, A.: DDTCDR: deep dual transfer cross domain recommendation. In: WSDM, pp. 331–339 (2020)
15. Liu, J., et al.: Exploiting aesthetic preference in deep cross networks for cross-domain recommendation. In: Proceedings of The Web Conference 2020, pp. 2768–2774 (2020)
16. Liu, M., Li, J., Li, G., Pan, P.: Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In: CIKM, pp. 885–894 (2020)
17. Ma, H., Yang, H., Lyu, M.R., King, I.: SoRec: social recommendation using probabilistic matrix factorization. In: CIKM, pp. 931–940 (2008)
18. Ma, J., Zhou, C., Cui, P., Yang, H., Zhu, W.: Learning disentangled representations for recommendation. In: NeuIPS, vol. 32 (2019)
19. Ma, J., Zhou, C., Yang, H., Cui, P., Wang, X., Zhu, W.: Disentangled self-supervision in sequential recommenders. In: KDD, pp. 483–491 (2020)
20. Man, T., Shen, H., Jin, X., Cheng, X.: Cross-domain recommendation: an embedding and mapping approach. In: IJCAI, vol. 17, pp. 2464–2470 (2017)

21. Oord, A.V.D., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
22. Qiu, R., Huang, Z., Yin, H.: Memory augmented multi-instance contrastive predictive coding for sequential recommendation. In: ICDM, pp. 519–528 (2021)
23. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: KDD, pp. 650–658 (2008)
24. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. Stat **1050**, 20 (2017)
25. Wang, K., Zhu, Y., Liu, H., Zang, T., Wang, C., Liu, K.: Inter-and intra-domain relation-aware heterogeneous graph convolutional networks for cross-domain recommendation. In: DASFAA, pp. 53–68 (2022)
26. Wang, X., Jin, H., Zhang, A., He, X., Xu, T., Chua, T.S.: Disentangled graph collaborative filtering. In: SIGIR, pp. 1001–1010 (2020)
27. Wang, Y., Tang, S., Lei, Y., Song, W., Wang, S., Zhang, M.: DisenHAN: disentangled heterogeneous graph attention network for recommendation. In: CIKM, pp. 1605–1614 (2020)
28. Wu, J., et al.: Self-supervised graph learning for recommendation. In: SIGIR, pp. 726–735 (2021)
29. Wu, J., et al.: DisenKGAT: knowledge graph embedding with disentangled graph attention network. In: CIKM, pp. 2140–2149 (2021)
30. Xia, X., Yin, H., Yu, J., Wang, Q., Cui, L., Zhang, X.: Self-supervised hypergraph convolutional networks for session-based recommendation. In: AAAI, vol. 35, pp. 4503–4511 (2021)
31. Xin, X., Liu, Z., Lin, C.Y., Huang, H., Wei, X., Guo, P.: Cross-domain collaborative filtering with review text. In: IJCAI (2015)
32. Xu, K., Xie, Y., Chen, L., Zheng, Z.: Expanding relationship for cross domain recommendation. In: CIKM, pp. 2251–2260 (2021)
33. Yan, H., Chen, X., Gao, C., Li, Y., Jin, D.: DeepAPF: deep attentive probabilistic factorization for multi-site video recommendation. TC **2**(130), 17–883 (2019)
34. Yu, W., Lin, X., Ge, J., Ou, W., Qin, Z.: Semi-supervised collaborative filtering by text-enhanced domain adaptation. In: KDD, pp. 2136–2144 (2020)
35. Zang, T., Zhu, Y., Liu, H., Zhang, R., Yu, J.: A survey on cross-domain recommendation: taxonomies, methods, and future directions. ACM Trans. Inf. Syst. **41**(2), 1–39 (2022)
36. Zang, T., Zhu, Y., Zhu, J., Xu, Y., Liu, H.: MPAN: multi-parallel attention network for session-based recommendation. Neurocomputing **471**, 230–241 (2022)
37. Zhao, C., Li, C., Fu, C.: Cross-domain recommendation via preference propagation GraphNet. In: CIKM, pp. 2165–2168 (2019)
38. Zhao, X., Yang, N., Yu, P.S.: Multi-sparse-domain collaborative recommendation via enhanced comprehensive aspect preference learning. In: WSDM, pp. 1452–1460 (2022)
39. Zheng, Y., et al.: Disentangling long and short-term interests for recommendation. In: Proceedings of the ACM Web Conference 2022, pp. 2256–2267 (2022)
40. Zhu, F., Chen, C., Wang, Y., Liu, G., Zheng, X.: DTCDR: a framework for dual-target cross-domain recommendation. In: CIKM, pp. 1533–1542 (2019)
41. Zhu, F., Wang, Y., Chen, C., Liu, G., Zheng, X.: A graphical and attentional framework for dual-target cross-domain recommendation. In: IJCAI, pp. 3001–3008 (2020)
42. Zhu, F., Wang, Y., Chen, C., Zhou, J., Li, L., Liu, G.: Cross-domain recommendation: challenges, progress, and prospects. arXiv preprint arXiv:2103.01696 (2021)