# Towards Universal Cross-Domain Recommendation

Jiangxia Cao
Institute of Information Engineering, CAS. School of Cyber Security, UCAS
caojiangxia@iie.ac.cn

Shaoshuai Li
MYbank, Ant Group
lishaoshuai.lss@mybank.cn

Bowen Yu
DAMO Academy, Alibaba Group
yubowen.ybw@alibaba-inc.com

Xiaobo Guo
MYbank, Ant Group
jefflittleguo.gxb@mybank.cn

Tingwen Liu★
Institute of Information Engineering, CAS. School of Cyber Security, UCAS
liutingwen@iie.ac.cn

Bin Wang
Xiaomi AI Lab, Xiaomi Inc
wangbin11@xiaomi.com

## ABSTRACT

In industry, web platforms such as Alibaba and Amazon often provide diverse services for users. Unsurprisingly, some developed services are data-rich, while some newly started services are data-scarce accompanied by severe data sparsity and cold-start problems. To alleviate the above problems and incubate new services easily, cross-domain recommendation (CDR) has attracted much attention from industrial and academic researchers. Generally, CDR aims to transfer rich user-item interaction information from related source domains (e.g., developed services) to boost recommendation quality of target domains (e.g., newly started services). For different scenarios, previous CDR methods can be roughly divided into two branches: (1) Data sparsity CDR fulfills user preference aided by other domain data to make intra-domain recommendations for users with few interactions, (2) Cold-start CDR projects user preference from other domain to make inter-domain recommendations for users with none interactions. In the past years, many outstanding CDR methods are emerged, however, to the best of our knowledge, none of them attempts to solve the two branches simultaneously. In this paper, we provide a unified framework, namely UniCDR, which can universally model different CDR scenarios by transferring the domain-shared information. Extensive experiments under the above 2 branches on 4 CDR scenarios and 6 public and large-scale industrial datasets demonstrate the effectiveness and universal ability of our UniCDR. Our source codes and a large-scale CDR dataset are released to facilitate academic research.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → **Neural networks**.

## KEYWORDS

Cross-Domain Recommendation; Collaborative Filtering; Multi-task Learning; Item Similarity Mining; Contrastive Learning

★Tingwen Liu is the corresponding author.



*WSDM '23, February 27-March 3, 2023, Singapore, Singapore*

ACM ISBN 978-1-4503-9407-9/23/02.
https://doi.org/10.1145/3539597.3570366

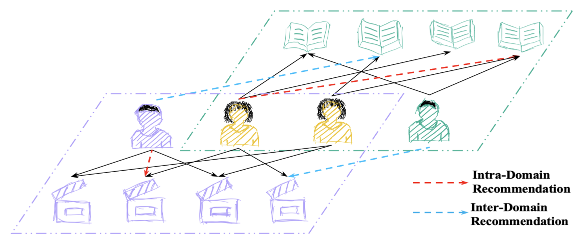


**Figure 1: Intra- and inter-domain recommendation of CDR.**

**ACM Reference Format:**
Jiangxia Cao, Shaoshuai Li, Bowen Yu, Xiaobo Guo, Tingwen Liu, Bin Wang. 2023. Towards Universal Cross-Domain Recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining (WSDM '23), February 27-March 3, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3539597.3570366

## 1 INTRODUCTION

In recent decades, Recommender Systems (RS) have experienced unprecedented evolution and played an indispensable role in effect billions of people's daily life. As a promising technique of RS, Collaborative Filtering (CF) provides a popular and widely used solution, from the early matrix decomposition [36, 39] to the latest deep neural network approaches [15, 45]. However, the CF-based methods always face two major challenges, the data sparsity [24, 39] and cold-start problem [50], which easily impede the user/item representation learning and severely influence recommendation accuracy. To alleviate such phenomenon, Cross-Domain Recommendation (CDR) was proposed [48]. The key idea is transferring the rich interaction information from the related source domain to improve prediction results on the target domain via overlapped users or overlapped items.

In the CDR family, for different research purposes, the previous works can be roughly divided into two branches while adopting a fundamental distinct learning framework. (1) For the data sparsity issue: this branch methods [26, 47] aim to transfer relative rich information from other domains to boost intra-domain recommendation performance for users with few interactions (e.g., recommend in-domain items for users just like the red dotted lines in Figure 1). Generally, several excellent works of this group follow the dual information transferring idea, such as CoNet [18] and Bi-TGCF [27]. They first encode user/item representations for each domain and then elaborate powerful information transferring modules to fuse

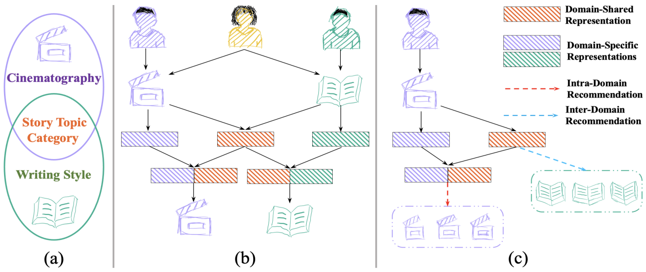


**Figure 2: The technical route for universal CDR.**

| Scenarios | Domains | | Overlapping | | Recommendation | |
|---|---|---|---|---|---|---|
| | Dual | Multi | User | Item | Intra | Inter |
| Scenario 1 | ✔ | | ✔ | | ✔ | |
| Scenario 2 | ✔ | | ✔ | | | ✔ |
| Scenario 3 | | ✔ | | ✔ | ✔ | |
| Scenario 4 | | ✔ | ✔ | | ✔ | |

**Table 1: The data format of four common CDR scenarios.**

and refine the learned representations for both domains. (2) For the cold-start issue: this branch methods aim to improve the more challenging inter-domain recommendation accuracy for new-coming users (e.g., the blue dotted lines in Figure 1). The recent efforts of this group are following the embedding and mapping idea, such as EMCDR [31] and SA-VAE [37]. In the training process, they first pre-train the user/item representations for each domain individually and then learn a mapping function to align the pre-trained representations according to the overlapped users. We thereby can map user representation from source to target domain and further recommend target domain items for source users.

In retrospect, the above technical frameworks achieve promising results in their branch, but they mainly focus on a narrow scenario and lack the capability to adapt to other scenarios. In this work, we rethink the cross-domain recommendation - toward a more comprehensive framework for various CDR scenarios. First of all, inspired by recent progresses [5, 6, 28], we also contemplate a challenging question of CDR: what information should be transferred to raise all domains' performance? In different domains, user behaviors always imply different preferences, and some trustworthy information could provide positive effects for other domains, while some biased preferences may lead to negative transferring problems. Here we give a toy example in Figure 2(a), there are two domains: "*Film*" and "*Book*" which contain several shared preferences and specific preferences. Intuitively, the shared user preferences such as "*Story Topic*" and "*Category*" are domain invariant, which means a user probably shows similar and stable taste in multiple domains for those preferences. Moreover, the specific user preferences such as "*Cinematography*" and "*Writing Style*" could provide precise intra-domain information which does not contribute to other domains. Therefore, the optimal way is to capture and transfer the most relevant factors (i.e., domain-shared information) to boost other domains, which sheds light on a reasonable technical route to build a universal cross-domain recommender system.

Motivated by the above observation, we propose UniCDR, a unified model that can transfer the domain-shared information in diverse CDR scenarios: (1) In the training stage, UniCDR learns the domain-shared and domain-specific representations for each user, as shown in Figure 2(b). (2) In the evaluation stage, UniCDR makes the intra- and inter- recommendations by using the domain-shared representation, as shown in Figure 2(c). In this work, we utilize some simple-yet-effective components to implement UniCDR, while each component has many alternative choices. Specifically, we first provide Mean, User-attention, and Item-similarity based aggregators to obtain the domain-shared and domain-specific representations. Then, we present Interaction and Domain masking mechanisms to produce augmented data to constrain the domain-shared representation encoding domain invariant information by our contrastive loss. Although such components are quite simple, we examine UniCDR under 4 different scenarios and 6 datasets and show that UniCDR achieves competitive performance to the latest baselines. Moreover, our model has great potential for industrial applications because of its universal ability and brief design style.

In summary, our major contributions are as follows:

- We highlight a universal perspective for CDR, which sheds light on building a new paradigm for various CDR scenarios by a single model. To our knowledge, this paper is the first work to model different CDR scenarios simultaneously.
- We propose a simple-yet-effective framework, UniCDR, which keeps a brief design style and shows competitive results to the elaborate state-of-the-art baselines.
- We perform experiments under 4 different CDR scenarios on 6 datasets and compare our model with 24 existing models to show its superior universality. Our codes and datasets are available at github[1] to facilitate the academic research.

## 2 PROBLEM STATEMENT

For a domain $\mathcal{D}$, we suppose that it consists of a user set $\mathcal{U}$, an item set $\mathcal{V}$ and an interaction set $\mathcal{E}$, i.e., $\mathcal{D} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ can be represented as a binary interaction matrix $\mathbf{A} \in \{0, 1\}^{|\mathcal{U}| \times |\mathcal{V}|}$. Let $\{\mathcal{D}^X, \mathcal{D}^Y, \mathcal{D}^Z\}$ denote the interaction data of domain $\{X, Y, Z\}$. CDR can be categorized from three aspects:

(1) **Dual** v.s. **Multi** domains: does it transfer knowledge over two or multiple domains?
(2) **User** v.s. **Item** overlapping: do users or items act as the overlapped role to bridge domains?
(3) **Intra** v.s. **Inter** recommendation: does it conduct intra- or inter-domain recommendation in the evaluation procedure?

From them, CDR community mainly focuses on 4 scenarios of the two branches in Table 1, and we consider them simultaneously.

## 3 UNIVERSAL CDR

Figure 3 sketches a high-level overview of our UniCDR under the dual domain setting, mainly including aggregators, mask mechanism and two kinds of loss functions. In this section, we first introduce the Embedding Layer, which gives several domain-specific and one domain-shared representation matrices for user/item. Second, we explain our Aggregator Architecture, an encoder with several implementation choices. Third, we present our Masking Mechanism

[1]https://github.com/cjx96/UniCDR

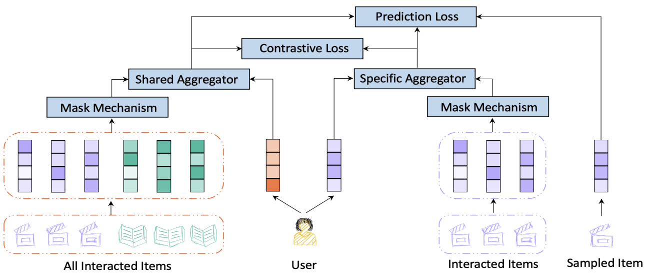


**Figure 3: The framework of UniCDR.**

and Contrastive Loss, which aim to model the correlation between domain-shared and domain-specific representations. Finally, we detail our model training procedure and the model evaluation procedure in different CDR scenarios. For better understanding, we express our methodology under the **Scenario 1** that two domains (e.g, domain $X$ and $Y$) have some overlapped users.

## 3.1 Embedding Layer

As discussed in Introduction, we conclude that the domain-shared information should be captured and transferred across domains. To consider the complex specific/shared information, we introduce $\mathbf{U}^X \in \mathbb{R}^{|\mathcal{U}^X|\times d}$, $\mathbf{U}^Y \in \mathbb{R}^{|\mathcal{U}^Y|\times d}$ and $\mathbf{U}^S \in \mathbb{R}^{|\mathcal{U}^X \cup \mathcal{U}^Y|\times d}$ to denote domain-$X$-specific, domain-$Y$-specific and domain-shared user representation matrices, where $d$ is the embedding dimension. Hence, for each user $u^X/u^Y$ in the domain $X/Y$, we can easily obtain the corresponding specific representation $\mathbf{u}^X/\mathbf{u}^Y$ and the shared representation $\mathbf{u}^S$ by a straightforward look-up operation. For items, we only use two domain-specific matrices $\mathbf{V}^X \in \mathbb{R}^{|\mathcal{V}^X|\times d}$, $\mathbf{V}^Y \in \mathbb{R}^{|\mathcal{V}^Y|\times d}$, since items are specific in the default user-overlapped setting[2].

## 3.2 Aggregator Architecture

As a well-explored backbone component of RS, several techniques (e.g., attention mechanism [43] and graph neural networks [45]) are successfully deployed to capture users personalized interaction information. We examined three candidate aggregators for UniCDR, but some other aggregators could also be adopted in different scenarios. Assume we have a user $u^X$ and his/her historically interacted items $\mathcal{H}_u^X = \{v_1^X, v_2^X, v_3^X, \dots\}$ and optional $\mathcal{H}_u^Y = \{v_1^Y, v_2^Y, v_3^Y, \dots\}$ (since some users are non-overlapped). For notation brevity, we omit the domain subscripts of user/item representations as the $\mathbf{u}/\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots\}$ in aggregator explanation.

*3.2.1 Mean-pooling aggregator.* We first utilize a direct way to aggregate the interacted items information equally, where we take the element-wise MEAN$(\cdot)$ of the item representations:

$$\mathbf{h} = \text{MEAN}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots\})\mathbf{W}_{\text{agg}}. \tag{1}$$

The $\mathbf{W}_{\text{agg}} \in \mathbb{R}^{d\times d}$ is a learnable parameter matrix and $\mathbf{h}$ is the final output. Actually, the above mean pooling aggregator is similar to the convolutional propagation rule in the GCN [22], such a simple design was widely deployed in industry, such as the PinSage [46] and YouTubeNet [8].

*3.2.2 User-attention-pooling aggregator.* Since the above simple pooling operation ignores the different item weights of different users, we further explore a more complex aggregator:

$$\begin{aligned} \boldsymbol{\alpha} &= \textsc{Softmax}(\{\alpha_1, \alpha_2, \alpha_3, \dots\}), \text{ where } \alpha_i = \textsc{Tanh}(\mathbf{W}_{\text{att}}\mathbf{v}_i + \mathbf{b})\mathbf{u}^\top, \\ \mathbf{h} &= \textsc{WeightedMean}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots\}, \boldsymbol{\alpha})\mathbf{W}_{\text{agg}}, \end{aligned} \tag{2}$$

where the normalized vector $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{H}_u|}$ denotes the non-negative user aware attention weights for interacted items $v \in \mathcal{H}_u$, and the $\textsc{Softmax}(\cdot)$ constraints that the attention weights sum is 1. Indeed, the above attention aggregator is closely related to GAT [44], by calculating the importance of neighbors for weighted propagation.

*3.2.3 Item-similarity-pooling aggregator.* Besides user-CF, item-CF also achieves great success in industry [1, 40], since users are always interested in coherent and similar items. Therefore, we are also curious about how to inject the item similarity information into UniCDR. In this work, inspired by the progress of the item-item linear model [34, 41], we provide a simple-yet-effective way to exploit the item similarity. Specifically, given the interaction matrix $\mathbf{A} \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{V}|}$, we first generate the item-item similarity weights $\mathbf{B} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ by EASE$^R$ [41]:

$$\underset{\mathbf{B}}{\text{argmin}}\ \|\mathbf{A} - \mathbf{A}\mathbf{B}\|_F^2 + \lambda_F\|\mathbf{B}\|_F^2 \quad \text{s.t. } \textsc{Diag}(\mathbf{B}) = 0. \tag{3}$$

The $\lambda_F$ is a hyperparameter of Frobenius norm regularizer, and the $\textsc{Diag}(\cdot)$ means the matrix diagonal elements (this constrain avoids a trivial case $\mathbf{B} = \mathbf{I}$). Mathematically, the above-constrained optimization problem has the following simple and desirable closed-form solution [41]:

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\big(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\big), \quad \text{where } \mathbf{P} = (\mathbf{A}^\top\mathbf{A} + \lambda_F\mathbf{I})^{-1}. \tag{4}$$

The $\textsc{DiagMat}(\cdot)$ denotes a diagonal matrix, $\mathbf{1}$ is a vector of ones and $\oslash$ means element-wise division. Based on the Eq.(4), we can easily conduct the calculating process by Algorithm 1. Afterward, we could leverage the pre-processed item-item similarity matrix $\mathbf{B}$ as prior knowledge[3] to guide our aggregator:

$$\begin{aligned} \boldsymbol{\alpha} &= \textsc{Normlize}(\{\alpha_1, \alpha_2, \alpha_3, \dots\}), \text{ where } \alpha_i = (\mathbf{A}\mathbf{B})_{u,v_i}, \\ \mathbf{h} &= \textsc{WeightedMean}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots\}, \boldsymbol{\alpha})\mathbf{W}_{\text{agg}}, \end{aligned} \tag{5}$$

where the $\mathbf{A}\mathbf{B} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{V}|}$ denote the personalized item score matrix. In retrospect, this aggregator builds a trade-off between GCN and GAT, that we generate an extra item similarity aware user-item score matrix to guide the convolutional propagation.

**Algorithm 1:** Item similarity pre-processing in Python

**Input:** Interaction matrix A and hyperparameter $\lambda_F$
**Output:** Item-item similarity matrix B

```
G = (A.T @ A).toarray() # The item-item coherent matrix
diag_indices = numpy.diag_indices_from(G)
G[diag_indices] += λF  # G + λF * I
P = numpy.linalg.inv(G) # The inverse matrix
B = P / (-numpy.diag(P))
B[diag_indices] = 0
```

[2]For item-overlapped Scenario 3, we further add an extra domain-shared item representation matrix $\mathbf{V}^S$ for our shared aggregator. Other than that there is no difference.

[3]To accelerate computation, we filter some smaller elements as zeros.

*3.2.4 Domain-specific & domain-shared representations.* Based on the above three aggregators, we can generate the corresponding domain-specific and domain-shared representation by giving different information. For instance, given a user $u^X$ and his/her historically interacted items $\mathcal{H}_u^X$ and optional $\mathcal{H}_u^Y$, we have:

$$\begin{aligned}\mathbf{h}_u^X &= \lambda_A \cdot \mathbf{u}^X + (1-\lambda_A)\cdot \textsc{Aggregator}^X(\mathbf{u}^X, \mathcal{H}_u^X),\\ \mathbf{h}_u^S &= \lambda_A \cdot \mathbf{u}^S + (1-\lambda_A)\cdot \textsc{Aggregator}^S(\mathbf{u}^S, \mathcal{H}_u^S),\end{aligned} \tag{6}$$

where $\mathcal{H}_u^S = \{\mathcal{H}_u^X, \mathcal{H}_u^Y\}$ means all interacted items, hyperparameter $\lambda_A$ controls the contribution rate between user and item information. Note that Eq.(6) also holds on domain $Y$, and we use three separate $\textsc{Aggregator}^X(\cdot)$, $\textsc{Aggregator}^Y(\cdot)$ and $\textsc{Aggregator}^S(\cdot)$ to generate the specific and shared representations $\mathbf{h}_u^X/\mathbf{h}_u^Y$ and $\mathbf{h}_u^S$.

## 3.3 Masking Mechanism and Contrastive Loss

Recently, contrastive learning (CL) achieves huge progress [6] in RS. It first generates multiple-view augmentation data by infusing perturbations on the original interaction data (e.g., masking mechanism) to form positive pairs. A series of unrelated instances are also sampled toward each original instance to form negative pairs. It then conducts a binary discriminator to identify the positive/negative sample pairs [16]. Actually, from the mutual information perspective, CL could maximize the mutual information [2] under different views (e.g., the shared-specific representation pairs in our setting). Therefore, we can further encourage our domain-shared aggregator to extract the domain-invariant information to optimize the likelihood of positive cases with multiple views of domain-specific information.

In the following, we first present our masking mechanism, and then express our contrastive loss calculating process.

*3.3.1 Masking mechanism.* To produce high-quality augmented data, we explore interaction masking and domain masking.

*Interaction masking.* We randomly remove a portion of interacted items in $\mathcal{H}_u$ to construct diverse item context information for the model to contrast with. To do this, we generate a masking vector $\boldsymbol{m} \in \{0,1\}^{|\mathcal{H}_u|}$ from Bernoulli distribution with probability $p$, where $m_i$ determines whether to remove the interacted items.

$$\widetilde{\mathcal{H}}_u = \textsc{Mask}(\mathcal{H}_u, \boldsymbol{m}), \text{ where } m_i = \textsc{Bernoulli}(p), \tag{7}$$

where $\widetilde{\mathcal{H}}_u$ denote the augmented data, and we leverage this mask to produce input $\widetilde{\mathcal{H}}_u^X, \widetilde{\mathcal{H}}_u^Y$ for our specific/shared aggregators.

*Domain masking.* Apart from interaction-level augmentation, we also consider a domain-level augmentation to encourage domain-shared representation encoding the domain-invariant information. Formally, we utilize the other domain interacted items to act as the shared aggregator input:

$$\widehat{\mathcal{H}}_{u,X}^S = \{\varnothing, \widetilde{\mathcal{H}}_u^Y\}, \quad \widehat{\mathcal{H}}_{u,Y}^S = \{\widetilde{\mathcal{H}}_u^X, \varnothing\} \tag{8}$$

where $\varnothing$ means empty item set, and $\widehat{\mathcal{H}}_{u,X}^S/\widehat{\mathcal{H}}_{u,Y}^S$ denotes the input for our shared aggregator in the training domain $X/Y$, respectively.

*3.3.2 Contrastive loss calculating process.* Based on the above two masking, we can first obtain the paired specific-shared augmentation representations $(\widetilde{\mathbf{h}}_u^X, \widehat{\mathbf{h}}_{u,X}^S)$ and $(\widetilde{\mathbf{h}}_u^Y, \widehat{\mathbf{h}}_{u,Y}^S)$ by Eq.(6). Then, we introduce a discriminator function to estimate the mutual information between the representation pairs. To be specific, taking domain $X$ as an example, we have (also holds in domain $Y$ as $\mathcal{L}_{\mathsf{con}}^Y$):

$$\mathcal{L}_{\mathsf{con}}^X = \sum_{u\in\mathcal{U}^X} [-\log \textsc{Disc}^X(\widetilde{\mathbf{h}}_u^X, \widehat{\mathbf{h}}_{u,X}^S) - \log(1-\textsc{Disc}^X(\widetilde{\mathbf{h}}_{\overline{u}}^X, \widehat{\mathbf{h}}_{u,X}^S))] \tag{9}$$

where the $\widetilde{\mathbf{h}}_{\overline{u}}^X$ denotes a negative specific representation by an arbitrary user (i.e., $\overline{u} \neq u$). The $\textsc{Disc}^X(\cdot) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the discriminator function in domain $X$ and we implement it by an bilinear function as follows:

$$\textsc{Disc}^X(\widetilde{\mathbf{h}}_u^X, \widehat{\mathbf{h}}_{u,X}^S) = \textsc{Sigmoid}\big(\widetilde{\mathbf{h}}_u^X \mathbf{W}_{\mathtt{disc}}^X (\widehat{\mathbf{h}}_{u,X}^S)^\top\big), \tag{10}$$

where $\mathbf{W}_{\mathtt{disc}}^X \in \mathbb{R}^{d\times d}$ is a parameter matrix. By optimizing $\mathcal{L}_{\mathsf{con}}^X$ and $\mathcal{L}_{\mathsf{con}}^Y$, the shared representation is encouraged to extract the domain-invariant information to maximize the likelihood.

## 3.4 Prediction Loss

In RS, two classes of loss functions are widely used: pairwise and pointwise loss. The pairwise loss is a ranking loss, which encourages the user to have a higher score for the interacted items than the items not interacted yet, such as Hinge [17] and BPR [36] loss. The pointwise loss aims to predict the user-item score precisely and model the real data distribution, such as the BCE [15] and Squared [7] loss. In this work, we follow previous works [3, 27] and denote the user-item interaction as an implicit score $\in \{0,1\}$, thus we exploit the BCE loss to make prediction (the pairwise loss will be explored in the future work). For domain $X$, we have:

$$\begin{aligned}\mathcal{L}_{\mathsf{pred}}^X &= \sum_{(u,v_i)\in\mathcal{E}^X} [-\log y_{(u,v_i)}^X - \log(1-y_{(u,\overline{v}_i)}^X)],\\ y_{(u,v_i)}^X &= \textsc{Sigmoid}\big(\textsc{Score}^X(\widetilde{\mathbf{h}}_u^X, \mathbf{v}_i^X) + \textsc{Score}^X(\widehat{\mathbf{h}}_{u,X}^S, \mathbf{v}_i^X)\big),\end{aligned} \tag{11}$$

where $y_{(u,v_i)}^X/y_{(u,\overline{v}_i)}^X$ denote the positive/negative interaction prediction scores. The $\textsc{Score}^X(\cdot)$ is a replaceable score function, we implement it as dot product for the sake of simplicity.

## 3.5 Training and Evaluation

*3.5.1 Model training.* Assuming we optimize our model in the first CDR scenario, its final training objective is as follows:

$$\mathcal{L} = \lambda(\mathcal{L}_{\mathsf{pred}}^X + \mathcal{L}_{\mathsf{pred}}^Y) + (1-\lambda)(\mathcal{L}_{\mathsf{con}}^X + \mathcal{L}_{\mathsf{con}}^Y). \tag{12}$$

The $\lambda$ is a harmonic factor to control the loss contribution. Besides, the above loss function can be easily extended to other CDR scenarios with same form, and we optimize it in a mini-batch manner.

*3.5.2 Model Evaluation.* In the evaluation process, we ignore the masking mechanism and generate the domain-specific and domain-shared representations $\mathbf{h}_u^X/\mathbf{h}_u^Y$ and $\mathbf{h}_u^S$ for user $u$ in domain $X/Y$. Afterward, for a user in domain $X$, we can predict the intra-domain item score:

$$\textsc{Sigmoid}\big(\textsc{Score}^X(\mathbf{h}_u^X, \mathbf{v}_i^X) + \textsc{Score}^X(\mathbf{h}_u^S, \mathbf{v}_i^X)\big), \tag{13}$$

or predict the inter-domain item score in domain $Y$:

$$\textsc{Sigmoid}\big(\textsc{Score}^Y(\mathbf{h}_u^S, \mathbf{v}_i^Y)\big). \tag{14}$$

Note that Eq.(13) and Eq.(14) also holds the user in domain $Y$.

**Table 2: Statistics of four CDR scenarios.**

| Scenarios | Datasets | $\lvert\mathcal{U}\rvert$ | $\lvert\mathcal{V}\rvert$ | Training | Valid | Test |
|---|---|---|---|---|---|---|
| Scenario 1 | Sport | 9,928 | 30,796 | 92,612 | - | 8,326 |
| | Cloth | 9,928 | 39,008 | 87,829 | - | 7,540 |
| | Elec | 3,325 | 17,709 | 50,407 | - | 2,559 |
| | Phone | 3,325 | 38,706 | 115,554 | - | 2,560 |
| Scenario 2 | Sport | 27,328 | 12,655 | 163,291 | 3,589 | 3,546 |
| | Cloth | 41,829 | 17,943 | 187,880 | 3,156 | 3,085 |
| | Game | 25,025 | 12,319 | 155,036 | 1,381 | 1,304 |
| | Video | 19,457 | 8,751 | 156,091 | 1,435 | 1,458 |
| Scenario 3 | M1 | 7,109 | 2,198 | 48,302 | 3,526 | 3,558 |
| | M2 | 2,697 | 1,357 | 19,615 | 1,362 | 1,310 |
| | M3 | 3,328 | 1,245 | 23,367 | 1,629 | 1,678 |
| | M4 | 5,482 | 2,917 | 41,226 | 2,720 | 2,727 |
| | M5 | 6,466 | 9,762 | 77,173 | 3,090 | 3,154 |
| Scenario 4 | D1 | 231,444 | 2,096 | 491,098 | 13,435 | 13,437 |
| | D2 | 507,715 | 595 | 1,068,490 | 36,013 | 35,985 |
| | D3 | 773,188 | 1,312 | 3,785,720 | 92,659 | 92,672 |

# 4 EXPERIMENTS

## 4.1 Datasets

Following many previous CDR works, we also conduct experiments under the public Amazon[4] dataset for a fair comparison. Specifically, for the dual-intra-domain recommendation **Scenario 1**, we select a series of preprocessed datasets by DisenCDR [5]: Sport&Cloth, Elec&Phone. For the dual-inter-domain recommendation **Scenario 2**, we utilize two preprocessed datasets by CDRIB [6]: Sport&Cloth, Game&Video. For the item-overlapped multi-intra-domain recommendation **Scenario 3**, we exploit the preprocessed dataset by M$^3$Rec [4], which includes five anonymized countries interaction data of electronics domain[5]. For the user-overlapped multi-intra-domain recommendation **Scenario 4**, due to the lack of public dataset in this scenario, we build a financial dataset that collects from three real services of MYbank. Compared with the E-commerce service of Amazon, the financial service of MYbank is special because the item set is much smaller. The detailed statistics of the datasets in four scenarios are summarized in Table 2.

## 4.2 Experimental Setting

*4.2.1 Evaluation Protocol.* Following previous works [3, 27, 50], we also leverage the leave-one-out technique to calculate the prediction results for the first three CDR scenarios. Specifically, as suggested by recent studies for unbiased evaluation [23], we randomly sample 999 negative items for each ground-truth item in the validation/testing set. For the last CDR scenario, we use the full-rank method [41] to calculate the prediction results. Then, we rank the items list and evaluate the $top-10$ items with the two widely-used metrics: NDCG (Normalized Discounted Cumulative Gain) and HR (Hit Ratio).

*4.2.2 Compared Baselines.* We compare UniCDR with the following four groups of state-of-the-art methods:

*Single-domain baselines.* We first compare with BPRMF [36], NeuMF [15], CML [17], EASE$^R$ [41], they are typically methods for industrial RS because of the direct idea and simple architecture. We also make comparison with the latest graph-based efforts: Random Walk [1], NGCF [45] and LightGCN [14], which aggregate the high-order neighbor information for user/item and achieve great progress for RS. For those methods, we mix all domain interaction data as one domain to train them.

*Intra-domain CDR baselines.* In this branch of CDR methods, we compare with CoNet [18], DDTCDR [26], they assign MLP-based encoder for each domain and devise different dual transferring modules between source and target domains. We also compare with the recent GNN based model, PPGN [47], Bi-TGCF [27] and DisenCDR [5], they utilize multiple neighboring information to enhance the user/item representations. Note that, the DisenCDR also aims at learning the domain-shared and domain-specific representations by variational inference framework [20].

*Inter-domain CDR baselines.* In this branch CDR methods, the EMCDR [31], SSCDR [19], TMCDR [50] and SA-VAE [37] mainly rely on the embedding-and-mapping paradigm: training an alignment function to project the user embeddings from source domain to target domain. Apart from them, CDRIB [6] proposes a new framework for this branch, which aims at learning an unbiased representation to encode the domain invariant information by the variational information bottleneck principle [42].

*Multi-domain baselines.* For the user overlapped scenario, GA-MTCDR [49] and HeroGraph [9] are two graph-based approaches by modeling the mixed global graph and local graphs simultaneously. The FOREC [3] and M$^3$Rec [4] are designed for item-overlapped scenario, while the M$^3$Rec focuses on the intra-domain and inter-domain item similarity mining. Moreover, the multi-task framework is also a promising way for multi-domain recommendation, we adapt the Cross-Stitch [33], MMoE [30] and STAR [38] into CDR by introducing MLP-based shared encoders and specific decoders.

*4.2.3 Implement Details.* To make fair comparisons with baseline methods, we follow their hyperparameter setting. The representation dimension $d$ is fixed as 128, the dropout rate is 0.3, the learning rate is set as 0.001, the mini-batch number is fixed as 1024, and the number of negative samples is selected from $\{1, 5, 10\}$. The Adam [21] optimizer is leveraged to update model parameters. Besides, for the specific parameters of our UniCDR, we suggest setting the Frobenius norm coefficient $\lambda_F$ in range 50~100 with step 5 in item similarity pre-processing stage, the factor $\lambda_A$ of our aggregator in range 0~1 at interval of 0.1, the probability $p$ of our interaction mask in range 0~0.5 at interval of 0.1, and the factor $\lambda$ of our training loss in range 0~1 at interval of 0.1. In next section, we report the UniCDR results under the User-attention aggregator for Scenario 1, Mean aggregator for Scenario 2, Item-similarity aggregator for Scenario 3, and Mean aggregator for Scenarios 4.

## 4.3 Performance Comparisons

Table 3, 4, 5 and 6 report the recommendation performances of all methods on the 4 different CDR scenarios, respectively. Note that we remove the domain masking and add an extra shared item

[4]http://jmcauley.ucsd.edu/data/amazon/index_2014.html
[5]We filter the redundant cold-start items in its valid/test datasets.

**Table 3: Performance comparison (%) of UniCDR(User-Att) on dual-intra-domain recommendations.**

| Datasets | Metric@10 | Single-Domain Methods | | | | Cross-Domain Methods | | | | | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BPRMF | NeuMF | NGCF | LightGCN | CoNet | DDTCDR | PPGN | Bi-TGCF | DisenCDR | UniCDR |
| Sport | HR | 10.43 | 10.74 | 13.13 | 13.19 | 12.09 | 11.86 | 15.10 | 14.83 | 17.55 | **18.37** |
| | NDCG | 5.41 | 5.46 | 6.87 | 6.94 | 6.41 | 6.37 | 8.03 | 7.95 | 9.46 | **10.98** |
| Cloth | HR | 11.53 | 11.18 | 13.22 | 13.58 | 12.40 | 12.54 | 14.23 | 14.68 | 16.31 | **17.85** |
| | NDCG | 6.25 | 6.02 | 6.97 | 7.29 | 6.62 | 7.13 | 7.68 | 7.93 | 9.03 | **11.20** |
| Elec | HR | 15.71 | 16.17 | 18.55 | 19.17 | 17.22 | 18.47 | 21.68 | 22.14 | **24.57** | 22.92 |
| | NDCG | 9.19 | 9.24 | 10.87 | 10.28 | 9.86 | 11.08 | 11.63 | 12.20 | **14.51** | 13.83 |
| Phone | HR | 16.32 | 15.84 | 22.79 | 23.25 | 17.66 | 17.23 | 24.54 | 25.71 | **28.76** | 24.72 |
| | NDCG | 8.53 | 8.02 | 12.38 | 12.72 | 9.30 | 8.58 | 13.34 | 13.93 | **16.13** | 13.77 |

**Boldface** and underlined numbers denote the best and runner-up results of all methods, respectively.

**Table 4: Performance comparison (%) of UniCDR(Mean) on dual-inter-domain recommendations.**

| Datasets | Metric@10 | Single-Domain Methods | | | Cross-Domain Methods | | | | | Ours |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CML | BPRMF | NGCF | EMCDR | SSCDR(CML) | TMCDR | SA-VAE | CDRIB | UniCDR |
| Sport | HR | 5.82 | 5.75 | 7.22 | 7.44 | 7.27 | 7.18 | 7.51 | **12.04** | 11.20 |
| | NDCG | 3.29 | 3.16 | 3.63 | 3.71 | 3.75 | 3.84 | 3.72 | 6.22 | **7.04** |
| Cloth | HR | 6.97 | 6.75 | 7.07 | 7.29 | 6.12 | 8.11 | 7.21 | 12.19 | **12.48** |
| | NDCG | 3.92 | 3.26 | 3.48 | 4.48 | 3.06 | 5.05 | 4.59 | 6.81 | **7.52** |
| Game | HR | 2.82 | 3.77 | 5.14 | 4.63 | 3.48 | 5.36 | 5.84 | 8.51 | **8.78** |
| | NDCG | 1.44 | 1.89 | 2.73 | 2.24 | 1.59 | 2.58 | 2.78 | 4.58 | **4.63** |
| Video | HR | 3.07 | 4.46 | 7.41 | 7.94 | 5.51 | 8.85 | 7.46 | **13.17** | 10.74 |
| | NDCG | 1.30 | 2.36 | 3.87 | 4.29 | 2.61 | 4.41 | 3.71 | **6.49** | 5.89 |

**Table 5: Performance comparison (%) of UniCDR(Item-Sim) on item-overlapped multi-intra-domain recommendation.**

| Datasets | Metric@10 | Single-Domain Methods | | | | Cross-Domain Methods | | | | | | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NeuMF | LightGCN | Random Walk | $EASE^R$ | Cross-Stitch | MMoE | Bi-TGCF | STAR | FOREC | $M^3$Rec | UniCDR |
| M1 | HR | 62.73 | 64.73 | 64.66 | 70.80 | 64.46 | 65.73 | 66.86 | 62.93 | 65.06 | **73.13** | 69.08 |
| | NDCG | 46.31 | 48.30 | 48.05 | 54.95 | 49.15 | 48.98 | 50.46 | 46.57 | 52.05 | 55.83 | **59.57** |
| M2 | HR | 55.60 | 52.13 | 50.20 | 57.40 | 54.06 | 56.26 | 53.46 | 54.89 | 58.42 | **60.86** | 58.01 |
| | NDCG | 34.84 | 32.70 | 31.10 | 37.92 | 36.65 | 38.71 | 33.43 | 35.25 | 40.03 | 40.04 | **47.52** |
| M3 | HR | 60.40 | 56.26 | 57.53 | 63.60 | 59.46 | 61.53 | 58.73 | 60.80 | 64.13 | **66.53** | 64.60 |
| | NDCG | 36.57 | 34.22 | 34.89 | 40.13 | 39.13 | 41.30 | 35.77 | 37.09 | 41.88 | 43.35 | **53.24** |
| M4 | HR | 40.33 | 41.14 | 39.20 | 45.13 | 38.93 | 38.60 | 42.26 | 40.20 | 41.60 | **48.46** | 47.52 |
| | NDCG | 29.83 | 31.03 | 30.08 | 36.63 | 29.16 | 30.16 | 32.76 | 29.84 | 33.52 | 37.98 | **42.54** |
| M5 | HR | 12.26 | 17.13 | 17.73 | 19.13 | 17.06 | 16.66 | 17.86 | 16.60 | 17.46 | **22.66** | 19.78 |
| | NDCG | 9.01 | 13.16 | 14.07 | 16.93 | 12.51 | 11.79 | 14.42 | 12.48 | 13.19 | **18.63** | 17.04 |

**Table 6: Performance comparison (%) of UniCDR(Mean) on user-overlapped multi-intra-domain recommendation.**

| Datasets | Metric@10 | Single-Domain Methods | | | | Cross-Domain Methods | | | | | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BPRMF | NeuMF | $EASE^R$ | LightGCN | MMoE | CoNet | Bi-TGCF | GA-MTCDR | HeroGraph | UniCDR |
| D1 | HR | 19.48 | 20.57 | 9.15 | 25.52 | 21.22 | 20.60 | 26.98 | 26.13 | 29.73 | **32.60** |
| | NDCG | 7.66 | 7.17 | 4.04 | 10.60 | 8.82 | 8.46 | 10.64 | 10.02 | 11.74 | **13.56** |
| D2 | HR | 50.45 | 52.92 | 50.07 | 56.18 | 56.22 | 53.53 | 60.48 | 59.59 | 61.49 | **64.37** |
| | NDCG | 33.50 | 35.73 | 28.53 | 37.09 | 38.63 | 37.66 | 47.19 | 47.67 | 49.57 | **50.48** |
| D3 | HR | 64.87 | 64.53 | 50.40 | 67.13 | 65.71 | 65.90 | 72.88 | 73.32 | 71.77 | **73.89** |
| | NDCG | 47.69 | 48.44 | 29.02 | 40.49 | 47.08 | 47.51 | 54.15 | 57.00 | 56.81 | **59.15** |

representation matrix in the item-overlapped Scenario 3. From them, we have the following insightful observations: (1) Compared to the single-domain baselines, the corresponding cross-domain baselines show consistent prediction improvement (i.e., MLP-based DDTCDR/NeuMF and graph-encoder-based Bi-TGCF/NGCF), which demonstrates that devising information transfer strategies across domain is better than directly training models in the mixed dataset. (2) The graph-encoder-based methods (i.e., PPGN, Bi-TGCF, and HeroGraph) reach significant improvement than other feed-forward neural network-based methods, such as CoNet and MMoE. This

**Table 7: Performance comparison (%) of aggregators.**

| Scenarios | Datasets | Metrics@10 | Aggregators | | |
|---|---|---|---|---|---|
| | | | Mean | User-Att | Item-Sim |
| Scenario 1 | Sport | HR | 14.26 | 18.37 | 16.96 |
| | | NDCG | 8.55 | 10.98 | 9.83 |
| | Cloth | HR | 15.23 | 17.85 | 16.32 |
| | | NDCG | 9.43 | 11.20 | 10.18 |
| | Elec | HR | 22.96 | 23.08 | 22.92 |
| | | NDCG | 13.45 | 13.73 | 13.83 |
| | Phone | HR | 22.86 | 22.38 | 24.72 |
| | | NDCG | 12.64 | 12.42 | 13.77 |
| Scenario 2 | Sport | HR | 11.20 | 7.57 | 9.88 |
| | | NDCG | 7.04 | 4.36 | 5.95 |
| | Cloth | HR | 12.48 | 9.25 | 10.89 |
| | | NDCG | 7.52 | 5.79 | 6.46 |
| | Game | HR | 8.78 | 6.37 | 7.44 |
| | | NDCG | 4.63 | 3.33 | 3.82 |
| | Video | HR | 10.74 | 5.71 | 10.38 |
| | | NDCG | 5.89 | 3.12 | 5.67 |

**Table 8: Performance comparison (%) of mask mechanism.**

| Scenarios | Datasets | Metrics@10 | Mask Mechanism | | |
|---|---|---|---|---|---|
| | | | $w/o$ Inter | $w/o$ Domain | UniCDR |
| Scenario 1 | Sport | HR | 17.03 | 18.55 | 18.37 |
| | | NDCG | 10.30 | 11.04 | 10.98 |
| | Cloth | HR | 17.21 | 17.15 | 17.85 |
| | | NDCG | 10.26 | 10.93 | 11.20 |
| | Elec | HR | 22.18 | 22.38 | 22.92 |
| | | NDCG | 11.15 | 13.80 | 13.83 |
| | Phone | HR | 22.74 | 25.86 | 24.72 |
| | | NDCG | 12.45 | 14.38 | 13.77 |
| Scenario 2 | Sport | HR | 10.36 | 7.50 | 11.20 |
| | | NDCG | 6.33 | 4.38 | 7.04 |
| | Cloth | HR | 11.97 | 7.71 | 12.48 |
| | | NDCG | 7.45 | 4.29 | 7.52 |
| | Game | HR | 7.96 | 6.80 | 8.78 |
| | | NDCG | 4.01 | 3.21 | 4.63 |
| | Video | HR | 9.87 | 6.55 | 10.74 |
| | | NDCG | 5.02 | 3.29 | 5.89 |

observation indicates that modeling the high-order information over the interaction graph is a promising way to enhance recommendation quality. (3) The DisenCDR and CDRIB aim to capture and transfer the domain-shared information across domains. And they outperform other baselines, which indicates that suitable transferring strategies significantly affect the prediction results. (4) For all CDR scenarios, our proposed method UniCDR shows competitive prediction results with the corresponding latest graph-encoder baselines, e.g., DisenCDR, CDRIB, M$^3$Rec and HeroGraph (Nevertheless, we do not stack the graph-encoder to capture high-order neighboring information in this work). This fact reveals that capturing and transferring the domain-shared information is vital for any CDR scenarios. Compared with DisenCDR and CDRIB, they implement the goal by following variational inference framework [20], but we achieve it by our masking mechanism and contrastive loss.

## 4.4 Analysis of Aggregators

In this section, we explore the effect of our three different pooling aggregators on the intra-domain and inter-domain recommendations. Table 7 reports the prediction results of the first two CDR scenarios. Note that we keep the same hyper-parameter setting but only change the aggregator. From it, we can observe that: (1) The user-attention-pooling and item-similarity-pooling aggregators show significant improvement over the mean-pooling aggregator in the intra-domain recommendation setting, validating that considering the importance of different items is necessary for the intra-domain recommendation. (2) Nevertheless, the mean-pooling aggregator achieves a more robust recommendation result than other aggregators in the inter-domain recommendation setting. This observation indicates that an appropriate aggregator is needed for different CDR scenarios. For instance, our item-similarity-pooling aggregator shows the best result in Scenario 3, i.e., item-overlapped multi-intra-domain recommendation.

## 4.5 Analysis of Masking Mechanism

In this section, to verify that our presented mask mechanism can augment the learned representations for CDR, we conduct another analysis on the intra-domain and inter-domain recommendation (The prediction results are shown in Table 8). Specifically, the $w/o$ Inter and $w/o$ Domain are the degenerate model variants of our UniCDR without the interaction mask and domain mask, respectively. According to Table 8, we can draw the following conclusions: (1) Compared with the $w/o$ Inter variant, our UniCDR shows satisfying improvement, which indicates that the interaction mask is a reliable way to boost our model to encode the domain-shared information in all CDR scenarios. (2) Compared with the $w/o$ Domain variant, we find an interesting phenomenon that our UniCDR show comparable results in the intra-domain recommendation but reach superior improvements in the inter-domain recommendation. We attribute the reason that the domain mask could simulate the inter-domain recommendation in the training process for each user, which could further enhance the domain generalization ability for the learned domain-shared representations to alleviate the cold-start problem. In summary, the interaction mask can efficiently achieve better performance in universal CDR scenarios in our experiments. Meanwhile, the domain mask plays an essential role in the inter-domain recommendation but does not contribute significantly to the intra-domain recommendation.

## 4.6 Hyperparameter Discussion

This section investigates the hyperparameter influence of the aggregator factor $\lambda_A$ and the training loss factor $\lambda$ on the "Sport-Cloth" dataset of Scenario 1. **For the aggregator factor** $\lambda_A$, Figure 4(a) and Figure 4(b) describe the recommendation performance line charts of Sport and Cloth domain, respectively. From them, we can find that our model shows higher and more robust results when the $\lambda_A \in [0.4, 0.6]$. According to the trajectory of the data points, we think fusing the user information and his/her aggregated item interaction information is crucial to generate better representations. In another hand, $\lambda_A$ also builds a trade-off between the user-CF and item-CF, thus we suggest that tuning this factor to around 0.5 is a reasonable way to achieve great performance. **For the training loss factor** $\lambda$, the Figure 5(a) and Figure 5(b) depict its results on Sport and Cloth domain, respectively. As shown in Figure 5, the line charts show an uptrend and then a downtrend for $\lambda$. Compared with $\lambda = 0.1$, our model obtains steady improvements in the cases $\lambda \in [0.2, 0.4]$, which proves that our contrastive loss is helpful to capture the domain-shared information. Further, our model always

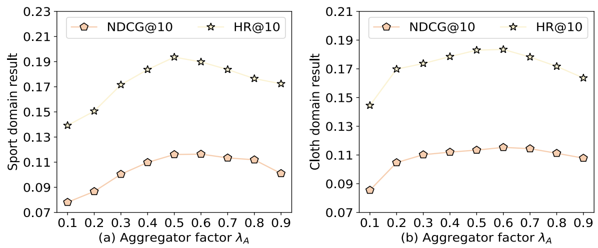


**Figure 4: Impact of the aggregator factor $\lambda_A$.**

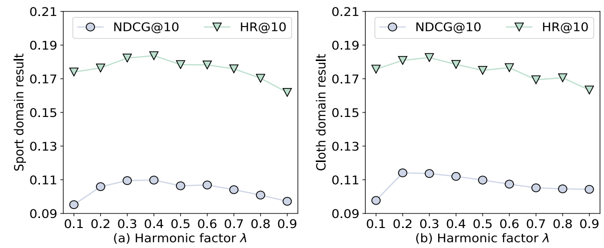


**Figure 5: Effect of the training loss factor $\lambda$.**

shows better prediction performances than other baselines in Table 3, which indicates that our UniCDR is robust to change the $\lambda$. Additionally, we observe that a larger $\lambda$ could achieve a faster speed for training convergence. That is, setting $\lambda$ as 0.3 or 0.4 might be a balanced choice between effectiveness and efficiency.

## 5 RELATED WORKS

In this section, we briefly survey the literature on recent CDR efforts developed for two different research goals:

*CDR for Intra-Domain Recommendation.* To alleviate the long-standing data sparsity problem and make intra-domain recommendation better, most CDR methods aim to integrate knowledge from multiple related domains via overlapped users or items. In early years, several works [12, 29] leverage a two-step migration idea: first pre-process user/item clusters and then learn a cluster-level "codebook" matrix [25] to transfer the rating pattern knowledge. Later, CMF [39] and CDCF [24] introduce a shared user matrix for all domains, and adapt the matrix decomposition algorithm to learn multiple rating matrices simultaneously. With the wave of neural networks, many deep models have been proposed to enhance knowledge transfer in recent years. The pioneering works are the MLP-based CoNet [18], DDTCDR [26] and graph-encoder-based PPGN [47], Bi-TGCF [27]. They mainly focus on designing robust information transferring modules to fuse and refine the learned representations for all domains, e.g., cross-connections network, user orthogonal mapping function, etc. In addition, the recent progress also points out the domain invariant information should be carefully considered (e.g., VDEA [28] and DisenCDR [5]), and they follow the variational auto-encoder framework [20] to reach the goal. On top of that, some other multi-domain scenarios in the industry are widely explored, such as user-overlapped and item-overlapped scenarios. GA-MTCDR [49], HeroGraph [9], FOREC [3] and $M^3$Rec [4] were proposed by mining user preference or item similarity.

*CDR for Inter-Domain Recommendation.* The cold-start inter-domain recommendation is a more challenging problem in RS, seriously affecting the new-coming user experience. With the help of other source domain interaction data, CDR provides a reliable way to mitigate the problem, attracting a surge of research interest. For example, CBMF [32] first preprocesses the user/item clusters and then learns a cluster-level matrix to describe the cross-domain user preferences. CST [35] transfers the pre-trained source domain user representations to initialize the user representations in the target domain, MVDNN [10] and GCBAN [13] transfer the users and item side information to alleviate the problem, respectively. Aside from that, EMCDR [31] proposes a pipeline style embedding-and-mapping framework: (1) first pre-training user embedding of source/target domain, (2) learning a mapping function with the overlapped users, (3) predicting target domain items for source users. Such framework motivates many researchers, SSCDR [19], TMCDR [50], PTUPCDR [51] and SA-VAE [37] were proposed by extending its mapping stage, e.g., introducing graph neural network, employing a user personalized meta network [11].

In summary, the previous CDR methods mainly focus on a single specific CDR scenario and their technical frameworks limit them to conduct in other scenarios to achieve expected results (e.g., using embedding-and-mapping framework to make intra-domain recommendations). Compared with them, UniCDR has important design differences that we consider the key challenge of CDR to transfer domain-shared information and devise a unified framework to model the intra- and inter-domain scenario at same time.

## 6 CONCLUSIONS

This paper proposes UniCDR, a flexible framework for universal CDR scenarios from the domain-shared information transferring perspective. We first point out that the optimal solution for all CDR scenarios is to capture and transfer the most relevant domain-shared information across domains. To implement the idea, our UniCDR keeps a simple and brief design style, to generate the domain-shared and domain-specific representation. Furthermore, to enhance the domain-shared representation, we introduce the masking mechanism and contrastive learning to constrain the shared representation encoding the unbiased domain invariant information. We conduct extensive experiments on 4 different CDR scenarios and 6 datasets, which demonstrate that our UniCDR shows universal ability and achieves competitive results with the latest state-of-the-art methods. Besides, considering the simple model architecture, our UniCDR has great potential for industrial applications. In the future, we will explore a graph-encoder-based aggregator and a more comprehensive masking mechanism for UniCDR.

## ACKNOWLEDGEMENT

This work was supported by the National Key Research and Development Program of China under Grant No.2021YFB3100600, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No.XDC02040400, the National Natural Science Foundation of China under Grant No.61902394, and the Youth Innovation Promotion Association of CAS under Grant No.2021153.

## REFERENCES


[1] Oren Barkan and Noam Koenigstein. 2016. Item2vec: neural item embedding for collaborative filtering. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*.
[2] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. 2018. Mutual Information Neural Estimation. In *International Conference on Machine Learning (ICML)*.
[3] Hamed Bonab, Mohammad Aliannejadi, Ali Vardasbi, Evangelos Kanoulas, and James Allan. 2021. Cross-Market Product Recommendation. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[4] Jiangxia Cao, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Item Similarity Mining for Multi-Market Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[5] Jiangxia Cao, Xixun Lin, Xin Cong, Jing Ya, Tingwen Liu, and Bin Wang. 2022. DisenCDR: Learning Disentangled Representations for Cross-Domain Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[6] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-Domain Recommendation to Cold-Start Users via Variational Information Bottleneck. In *IEEE International Conference on Data Engineering (ICDE)*.
[7] Chong Chen, Min Zhang, Yongfeng Zhang, Yiqun Liu, and Shaoping Ma. 2020. Efficient neural matrix factorization without sampling for recommendation. *ACM Transactions on Information Systems (TOIS)* (2020).
[8] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *ACM International Conference on Recommender Systems (RecSys)*.
[9] Qiang Cui, Tao Wei, Yafeng Zhang, and Qing Zhang. 2020. HeroGRAPH: A Heterogeneous Graph Framework for Multi-Target Cross-Domain Recommendation.. In *RecSys ORSUM Workshop*.
[10] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *The World Wide Web Conference (WWW)*.
[11] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning (ICML)*.
[12] Sheng Gao, Hao Luo, Da Chen, Shantao Li, Patrick Gallinari, and Jun Guo. 2013. Cross-domain recommendation via cluster-level latent factor model. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*.
[13] Jia He, Rui Liu, Fuzhen Zhuang, Fen Lin, Cheng Niu, and Qing He. 2018. A general cross-domain recommendation framework via Bayesian neural network. In *IEEE International Conference on Data Mining (ICDM)*.
[14] Xiangnan He, Kuan Deng, Xiang Wang, Yaliang Li, Yongdong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[15] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *The World Wide Web Conference (WWW)*.
[16] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. 2019. Learning deep representations by mutual information estimation and maximization. In *International Conference on Learning Representations (ICLR)*.
[17] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In *The World Wide Web Conference (WWW)*.
[18] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. Conet: Collaborative Cross Networks for Cross-Domain Recommendation. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[19] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-Supervised Learning for Cross-Domain Recommendation to Cold-Start Users. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[20] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations (ICLR)*.
[21] P. Diederik Kingma and Lei Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*.
[22] Thomas N Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
[23] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[24] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. In *International Joint Conference on Artificial Intelligence (IJCAI)*.
[25] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *International Joint Conference on Artificial Intelligence (IJCAI)*.
[26] Pan Li and Alexander Tuzhilin. 2020. DDTCDR: Deep Dual Transfer Cross Domain Recommendation. In *ACM International Conference on Web Search and Data Mining (WSDM)*.
[27] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[28] Weiming Liu, Xiaolin Zheng, Jiajie Su, Mengling Hu, Yanchao Tan, and Chaochao Chen. 2022. Exploiting Variational Domain-Invariant User Embedding for Partially Overlapped Cross Domain Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[29] Zhongqi Lu, Erheng Zhong, Lili Zhao, Evan Wei Xiang, Weike Pan, and Qiang Yang. 2013. Selective transfer learning for cross domain recommendation. In *SIAM International Conference on Data Mining (SDM)*.
[30] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[31] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach.. In *International Joint Conference on Artificial Intelligence (IJCAI)*.
[32] Nima Mirbakhsh and Charles X Ling. 2015. Improving Top-n Recommendation for Cold-Start Users via Cross-Domain Information. *ACM Transactions on Knowledge Discovery from Data (TKDD)* (2015).
[33] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. 2016. Cross-stitch networks for multi-task learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[34] Xia Ning and George Karypis. 2011. Slim: Sparse linear methods for top-n recommender systems. In *IEEE International Conference on Data Mining (ICDM)*.
[35] Weike Pan, Evan Xiang, Nathan Liu, and Qiang Yang. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI Conference on Artificial Intelligence (AAAI)*.
[36] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. *International Conference on Uncertainty in Artificial Intelligence (UAI)* (2009).
[37] Aghiles Salah, Thanh Binh Tran, and Hady Lauw. 2021. Towards Source-Aligned Variational Models for Cross-Domain Recommendation. In *ACM Conference on Recommender Systems (RecSys)*.
[38] Xiang-Rong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, et al. 2021. One Model to Serve All: Star Topology Adaptive Recommender for Multi-Domain CTR Prediction. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[39] Ajit P Singh and Geoffrey J Gordon. 2008. Relational Learning via Collective Matrix Factorization. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[40] Brent Smith and Greg Linden. 2017. Two decades of recommender systems at Amazon. com. *IEEE International Computing* (2017).
[41] Harald Steck. 2019. Embarrassingly shallow autoencoders for sparse data. In *The World Wide Web Conference (WWW)*.
[42] Naftali Tishby and Noga Zaslavsky. 2015. Deep Learning and The Information Bottleneck Principle. In *IEEE Information Theory Workshop (ITW)*.
[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.
[44] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations (ICLR)*.
[45] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[46] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[47] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-Domain Recommendation via Preference Propagation GraphNet. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[48] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. *ArXiv* (2021).
[49] Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu. 2021. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* (2021).
[50] Yongchun Zhu, Kaikai Ge, Fuzhen Zhuang, Ruobing Xie, Dongbo Xi, Xu Zhang, Leyu Lin, and Qing He. 2021. Transfer-Meta Framework for Cross-domain Recommendation to Cold-Start Users. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[51] Yongchun Zhu, Zhenwei Tang, Yudan Liu, Fuzhen Zhuang, Ruobing Xie, Xu Zhang, Leyu Lin, and Qing He. 2021. Personalized Transfer of User Preferences for Cross-domain Recommendation. *ArXiv* (2021).