Knowledge-Based Systems 310 (2025) 112919

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

TJMN: Target-enhanced joint meta network with contrastive learning for cross-domain recommendation

Xiaofei Zhu [a],*, Lele Duan [a], Stefan Dietze [b,c], Ran Yu [d]

[a] College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China
[b] Knowledge Technologies for the Social Sciences, Leibniz Institute for the Social Sciences, Cologne 50667, Germany
[c] Institute of Computer Science, Heinrich-Heine-University, Düsseldorf 40225, Germany
[d] Institute of Computer Science III, University of Bonn, Bonn 53115, Germany

ARTICLE INFO

Keywords:
Recommender systems
Cross-domain recommendation
Meta network
Contrastive learning

ABSTRACT

Cross-domain recommendation (CDR) provides a promising solution to mitigate the sparsity issue in the target domain by exploiting auxiliary information from the source domain. Recently, meta learning based methods have been proposed and achieved the state-of-the-art performance. However, these methods learn the transfer bridge solely relying on the source domain while the rich information from the target domain are ignored. Moreover, they leverage either a common transfer bridge or a personalized transfer bridge to transform user representations, without considering the multi-grained characteristics of user preference. In this paper, we propose a target-enhanced joint meta network with contrastive learning (JTMN) for cross-domain recommendation. To be specific, we develop a target bridge to incorporate information from the target domain to guide the learning process of user preference transfer. In addition, we introduce multi-grained transfer bridges to model the complex transfer patterns of user preference across different domains. At last, a target-aware contrastive learning layer is designed to obtain better user representations. The experimental results on six CDR tasks demonstrate that our proposed TJMN model significantly outperforms all strong baselines, especially when the training data become more sparse.

1. Introduction

In recent years, recommender systems have attracted a great amount of attention and emerged as a powerful strategy for alleviating the information overload problem [1–4]. Conventional research works mainly focus on leveraging users' historical interactions and have achieved encouraging performance. However, in the real-world scenarios, the historical interactions of users would be sparse or even unavailable, i.e., cold-start users, which leads to unsatisfying recommendation. To deal with this issue, cross-domain recommendation (CDR) [5–7], which aims to transfer the information from the source domain to the target domain, have been proposed as a promising solution. The core assumption of CDR is that the source domain and the target domain are relevant and knowledge learned from the source domain could be utilized to enhance the recommendation in the target domain. Existing methods usually rely on the embedding and mapping structure [8–10], which aim at learning a mapping function from source domain to target domain by leveraging the overlapping users. The main limitation of these methods is that they may be biased to the limited overlapping users and suffer from unsatisfying generalization ability.

More recently, some research works attempt to apply meta learning [11–13] to handle the above-mentioned issue due to its good generalization ability. Zhu et al. [14] propose a transfer-meta framework for CDR (TMCDR) by learning a common bridge between the source domain and the target domain. TMCDR is comprised of a transfer stage and a meta stage, where the former is utilized to obtain the pre-trained source and target models, and the latter is developed to implicitly transfer the embeddings from the source domain to the target domain. Since the relationships between user preference of the two domains may vary considerably, Zhu et al. [2] further propose a personalized transfer bridge of user preference for cross-domain recommendation (PTUPCDR). Instead of learning a common bridge for all users, PTUPCDR utilizes the characteristic embedding of each user in the source domain as input to generate the corresponding personalized bridge, i.e., the learned bridge will vary from user to user. With this strategy, the complex relationships between user preference of both domains can be well modeled in a fine-grained manner.

Although existing meta learning based methods have achieved encouraging performance, they still suffer from some defects. First, these

* Corresponding author.
E-mail addresses: zxf@cqut.edu.cn (X. Zhu), dll@2020.cqut.edu.cn (L. Duan), stefan.dietze@gesis.org (S. Dietze), ran.yu@uni-bonn.de (R. Yu).

https://doi.org/10.1016/j.knosys.2024.112919
Received 24 February 2023; Received in revised form 1 October 2024; Accepted 21 December 2024
Available online 30 December 2024
0950-7051/

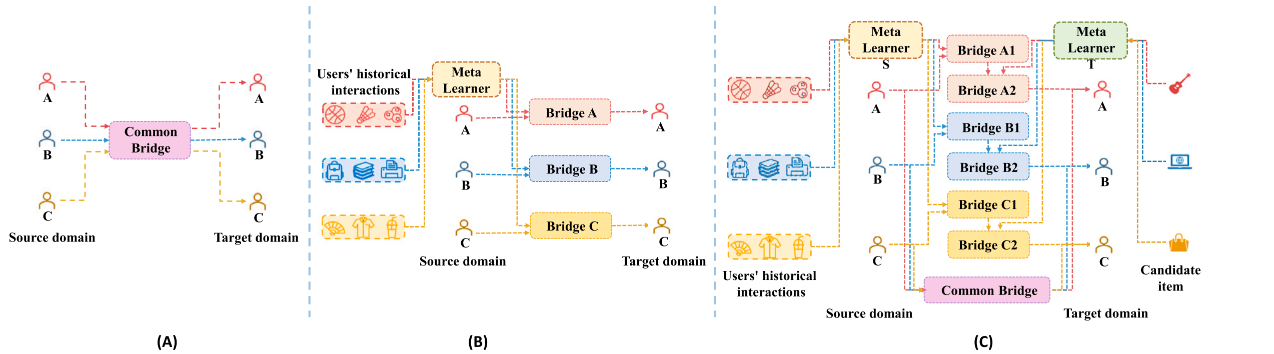

Fig. 1. The difference between our proposed approach TJMN and conventional meta network based methods. (A) All users share a common bridge. (B) A meta network is utilized to generate a personalized bridge for each user. (C) Our proposed approach TJMN leverages multi-granularity bridges to capture different user preference transfer patterns with distinct granularities.

methods learn the transfer bridge solely relying on the source domain while the information from the target domain is ignored. Since the information from the target domain is also important for bridging user preference between the source and target domains, further incorporating the information from the target domain will be useful to guide the learning procedure. Second, existing methods leverage either a common bridge or a personalized bridge to transform users' representations. The former could be considered as a coarse-grained bridge which captures the common user preference within both source and target domains, while the latter could be considered as a fine-grained bridge which models the personalized user preference within the two domains. We argue that simultaneously utilizing both kinds of bridges will bring more benefits for learning better representations. Third, existing methods learn user representations by only utilizing positive interacted items, which indicates user preference, while overlook what they dislike, thus cannot capture the rich partial order relation information of user preference.

To deal with the issues mentioned-above, in this paper, we propose a novel approach named Target-enhanced Joint Meta Network with Contrastive Learning (JTMN) for cross-domain recommendation. Specifically, we develop a target bridge which introduces the signal from the target domain to guide the learning process of user preference transfer between the two domains. Moreover, instead of learning either a coarse-grained bridge (i.e., common bridge [14]) or a fine-grained bridge (i.e., personalized bridge [2]), we develop multi-granularity transfer bridges, which include a tri-kernel personalized bridge and a common bridge, in order to capture different user preference transfer patterns with distinct granularities. The tri-kernel personalized bridge leverages users' positive and negative interaction items by generating three intent kernels, i.e., the positive intent kernel, the negative intent kernel and the mix intern kernel. Finally, we also propose a target-aware contrastive learning layer to learn better user representations by enhancing the contrastive learning with signals from the target domain. The difference between our work and conventional meta network based methods is illustrated in Fig. 1.

We carry out extensive experiments on six different cross-domain recommendation tasks, and the results show that our proposed method outperforms all state-of-the-art baselines in terms of both metrics, i.e., the mean absolute error (MAE) and the root mean square error (RMSE). The main contributions are summarized as follows:

- We propose to incorporate the target signal to guide the process of user preference transfer between the source and target domains. To the best of our knowledge, our proposed TJMN is the first work to exploit the target signal with a target bridge to capture the complex relationships between the user preference of the source and target domain.
- We develop a target-enhanced joint meta network with contrastive learning to capture user transferable characteristics from different perspectives, in which the multi-granularity transfer bridges are developed to capture different user preference transfer patterns with distinct granularities.
- Extensive experiments demonstrate that our proposed model is superior to the state-of-the-art baselines. Further ablation studies also verify the effectiveness of each module in our model.

The rest of the paper is organized as follows. Section 2 gives a brief review of the related work. In Section 3, we describe the details of our proposed approach. The experimental results of our empirical studies are discussed in Section 4 and we conclude the paper in Section 5.

2. Related work

Recommender systems (RS) have emerged as an indispensable component for alleviating the information overload problem in real-world applications. Previous research works usually suffer from the cold-start problem [14–17], where few user–item interactions are available for users for making effective recommendation. To handle this issue, cross-domain recommendation (CDR) have received considerable attention in recent years. Early studies on CDR are mainly based on the collaborative filtering framework [5,18–20]. CMF [5] aims to simultaneously factor several matrices and share the user embeddings across different domains. Temporal-domain CF [18] employs the cross-domain CF framework [21] to connect different temporal domains via introducing the group-level rating matrix, which can be considered as an expected rating provided by a user prototype on an item prototype. The relatedness of user-group memberships between successive temporal domains will be modeled to capture the user-interest drift. CDTF [19] models the interactions between domain-specific user factors and item factors by exploring the triadic relation user–item–domain.

In recent years, many deep learning based methods have been proposed to address the sparsity and cold-start problems in CDR. EMCDR [8] investigates the CDR problem from an embedding and mapping perspective. It first learns user representations in different domains independently, and then trains a cross-domain mapping function by minimizing the distance between the actual target embeddings and the transferred embeddings in the target domain for these overlapping users. DCDCSR [9] extends EMCDR by generating benchmark factors and mapping the latent factors in the target domain to fit these benchmark factors. It considers fine-grained sparsity degrees of individual users and items to combine the latent factors learned from both the source and target domains. Since the number of overlapping users between the source and target domains would be very limited in many real applications, these EMCDR-based methods may suffer from the unsatisfying generalization ability on cold-start users in the target domain. To deal with this issue, SSCDR [22] attempts to capture the cross-domain relationship via learning the latent vectors of users and items in metric space. It utilizes a semi-supervised approach to train a cross-domain mapping function and infers the latent vectors of the cold-start users based on users' neighborhood information.

More recently, some meta learning [2,14] based methods have been proposed and obtained the SOTA performance for the task of cross-domain recommendation. TMCDR [14] applies a transfer-meta framework for CDR by replacing the training procedure of these EMCDR-based methods. It is comprised of a transfer stage and a meta stage. Differ from these EMCDR-based methods which train the embedding model solely on overlapping users, the transfer stage trains a unique model for each domain and takes the source and target models as pre-trained models. In the meta stage, it learns a task-oriented meta network to transform embeddings of the overlapping users from the source domain to the target domain. As the complex relationships between user preference in different domains would vary from user to user, learning a common bridge shared by all users would be improper to capture users' individual preference. To alleviate this issue, PTUPCDR [2] leverages a meta network to generate a personalized bridge for each user, in which the bridge function depends on each user's characteristics. After training, user embeddings in the source domain will be fed into the meta-generated personalized bridge functions to obtain the transformed user embeddings in the target domain. REMIT [23] proposes to extract users' multiple interests by exploiting external knowledge such as item category and brand, and aggregate transformed interests based on a reinforcement learning framework. The limitation of this method is that it heavily relies on external knowledge. To address this problem, MIMNet [24] develops a multi-interest meta network to decouple users' multiple interests, and generates multi-interest bridges to transfer user representations from the source domain to the target domain. The main differences between our proposed approach and the state-of-the-art methods are three folds. First, to the best of our knowledge, our proposed approach is the first work that exploiting the target signal with a target bridge to capture the complex relationships between the user preference of the source and target domain. Second, we simultaneously employ both coarse-grained and fine-grained bridges to capture the common user preference as well as the personalized user preference between the source and target domains. Third, we further model users' negative transferable characteristics based on their negative interaction items in order to capture the rich partial order relation information of user preference.

3. Preliminaries

In cross-domain recommendation, we have a source domain and a target domain. Each domain includes users, items, and interactions ($e.g.,\ ratings$) between users and items. Let $U = \{u_1, u_2, \dots, u_n\}$ denote the user set and $V = \{v_1, v_2, \dots, v_m\}$ denote the item set, where n and m are the number of users and items, respectively. Let R denote the rating matrix and $r_{ij} \in R$ represents the interaction between the user u_i and the item v_j. We denote the user set, the item set and the rating matrix in the source domain as U^s, V^s and R^s. Similarly, we use U^t, V^t and R^t for the target domain. We denote the overlapping users between the two domains as $U^o = U^s \cap U^t$, and the cold-start users (i.e., users exist in the source domain while not in the target domain) as $U^c = \{u | u \in U^s \wedge u \notin U^t\}$. It is worth noting that, in cross-domain recommendation, the two domains have no shared items, i.e., $V^s \cap V^t = \emptyset$.

For each user u_i^s in source domain, we denote $S_{u_i}^p = \{v_1^{s,p}, v_2^{s,p}, \dots, v_{p_i}^{s,p}\}$ and $S_{u_i}^n = \{v_1^{s,n}, v_2^{s,n}, \dots, v_{n_i}^{s,n}\}$ as his corresponding positive and negative interaction items respectively, where p_i (n_i) denotes the number of positive (negative) interaction items, and $v_k^{s,p}$ ($v_k^{s,n}$) denotes the kth positive (negative) interaction item. We can transform the users and items into dense vectors, also called embeddings, with the latent factor model [25]. In this paper, we use $\mathbf{u}_i^* \in \mathbb{R}^d$ and $\mathbf{v}_i^* \in \mathbb{R}^d$ to denote the embeddings of the user u_i^* and the item v_i^* respectively, where d denotes the dimensionality of the embedding and $* \in \{s, t\}$ represents the label of domain. Specifically, we use $\mathbf{S}_{u_i}^p = \{\mathbf{v}_1^{s,p}, \mathbf{v}_2^{s,p}, \dots, \mathbf{v}_{p_i}^{s,p}\}$ to denote the embeddings of positive interaction items of user u_i^s in the source domain, where $\mathbf{v}_k^{s,p} \in \mathbb{R}^d$ is the embedding of the positive interaction item $v_k^{s,p}$. Similarly, we use $\mathbf{S}_{u_i}^n = \{\mathbf{v}_1^{s,n}, \mathbf{v}_2^{s,n}, \dots, \mathbf{v}_{n_i}^{s,n}\}$ to denote the embeddings of negative interaction items of user u_i^s in the source domain, where $\mathbf{v}_k^{s,n} \in \mathbb{R}^d$ is the embedding of the negative interaction item $v_k^{s,n}$.

Positive Characteristic Encoder. Inspired by the attention mechanism [26,27], the attention weights corresponding to different items could be leveraged to reflect their importance for personalized bridge functions. Therefore, the representation of the current user u_i^s characterized by the positive interaction items can be defined as follows:

$$\mathbf{p}_{u_i} = \sum_{k=1}^{p_i} \alpha_k^p \mathbf{v}_k^{s,p}, \tag{1}$$

$$\alpha_k^p = \frac{\exp(a_k^p)}{\sum_{j=1}^{p_i} \exp(a_j^p)}, \tag{2}$$

$$a_j^p = Attn(\mathbf{v}_j^{s,p}; \theta_p), \tag{3}$$

where $Attn(\cdot)$ represents the attention network and θ_p represents its corresponding parameters.

Negative Characteristic Encoder. The representation of user u_i^s characterized by the negative interaction items can be estimated in a similar way as the positive characteristic encoder. However, different from these positive interaction items, the negative interaction items usually contain no or little user preference information. Therefore, we adopt the mean pooling to obtain the representation of user u_i^s characterized by the negative interaction items as:

$$\mathbf{n}_{u_i} = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{v}_k^{s,n}. \tag{4}$$

The notations we will use throughout this paper are summarized in Table 1.

4. Model

In this section, we introduce the proposed Target-enhanced Joint Meta Network (TJMN), which attempts to learn multiple transfer bridges with different granularities of user preference as well as involve the information from the target domain to guide the learning process. The overall structure of the proposed model TJMN is illustrated in Fig. 2, which mainly consists of five components: (1) Tri-Kernel Personalized Bridge; (2) Common Bridge; (3) Target Bridge; (4) Target-Aware Contrastive Learning Layer; (5) Prediction Layer.

4.1. Tri-Kernel Personalized Bridge (TKPB)

When capturing users' transferable characteristics, existing research works [2,14,28,29] mainly focus on considering positive interaction items (i.e., items with high rating scores) while largely overlook these negative interaction items (i.e., items with lower rating scores or no interactions). Due to the data sparsity, these positive interaction items would not fully reflect a user preference, e.g., they can only reflect what a user likes while not she dislikes. To address this issue, we propose to incorporate the negative interaction items to capture users' negative transferable characteristics. The rationale to involve these negative interaction items is that they usually contain rich partial order relation information, i.e., these positive interaction items are usually more informative as compared to these negative interaction items. To effectively capture users' transferable characteristics, we attempt to simultaneously leverage users' positive and negative interaction items, and develop a Tri-Kernel Personalized Bridge (TKPB). In particular, TKPB contains two meta networks (i.e., the source-drive pos-meta network and the source-drive neg-meta network), which are utilized to generate the positive and negative intent kernels, respectively. A

Table 1
Notations and explanations.

Notation	Description
U	User set
V	Item set
R	Rating matrix
r_{ij}	Interaction between the user u_i and the item v_j
U^o	Overlapping users between the source and target domains
U^c	Cold-start users
u_i^s	The ith user in the source domain
$S_{u_i}^p$	Corresponding positive interaction items of u_i^s
$S_{u_i}^n$	Corresponding negative interaction items of u_i^s
$p_i(n_i)$	Number of positive (negative) interaction items
$v_k^{s,p}(v_k^{s,n})$	The kth positive (negative) interaction item
$\mathbf{u}_i^s$	ID embedding of u_i^s in the source domain
$\mathbf{p}_{u_i}(\mathbf{n}_{u_i})$	Representation of u_i^s characterized by the positive (negative) interaction items in the source domain
$\mathbf{w}_{u_i}^p$	Parameters of the positive personalized bridge function
$\mathbf{w}_{u_i}^n$	Parameters of the negative personalized bridge function
$\mathbf{w}_{u_i}$	Parameters of the mixed personalized bridge function
$\mathbf{w}_{v_j^t}$	Parameters of the personalized target bridge function
$\mathbf{u}_i^{t_p}$	Transformed representation of $\mathbf{u}_i^s$ by the mixed personalized bridge function
$\mathbf{p}_{u_i}^{t_p}$	Transformed representation of $\mathbf{p}_{u_i}$ by the positive personalized bridge function
$\mathbf{n}_{u_i}^{t_p}$	Transformed representation of $\mathbf{n}_{u_i}$ by the negative personalized bridge function
$\mathbf{u}_i^{t_c}$	Transformed representation of $\mathbf{u}_i^s$ by the common bridge
$\mathbf{p}_{u_i}^{t_c}$	Transformed representation of $\mathbf{p}_{u_i}$ by the common bridge
$\mathbf{n}_{u_i}^{t_c}$	Transformed representation of $\mathbf{n}_{u_i}$ by common bridge
$\mathbf{u}_{i,v_j^t}^{t_p}(\mathbf{u}_{i,v_j^t}^{t_c})$	Transformed representation of $\mathbf{u}_i^{t_p}(\mathbf{u}_i^{t_c})$ by the personalized target bridge function
$\hat{\mathbf{u}}_i^t$	The final transformed representation of u_i^s in target domain
$\hat{r}_{ij}$	The predicted rating user u_i gives to item v_j in target domain
$\mathcal{R}_o^t$	The ratings of overlapping users in the target domain
τ	The temperature parameter
λ	The trade-off parameter

mixed intent kernel is also introduced to capture users' comprehensive transferable characteristics. Based on the three intent kernels, TKPB generates three corresponding personalized bridge functions, including the positive personalized bridge function, the negative personalized bridge function, and the mixed personalized bridge function.

4.1.1. Source-drive pos-meta network

Inspired by the meta-learning [2,30], we propose a meta network to generate the positive intent kernel based on users' positive transferable characteristics. It is worth noting that the input of this meta network solely relies on the positive interaction items in the source domain, thus we call the meta network as the source-drive pos-meta network. To be specific, the source-drive pos-meta network $g^p(\cdot)$ is defined as a two-layer feed-forward network, which takes $\mathbf{p}_{u_i}$ as input:

$$\tilde{\mathbf{w}}_{u_i}^p = g^p(\mathbf{p}_{u_i};\phi_p), \tag{5}$$

where ϕ_p are learnable parameters, $\tilde{\mathbf{w}}_{u_i}^p \in \mathbb{R}^{d^2}$ is a vector whose size depends on the structure of its corresponding personalized bridge function. We then reshape $\tilde{\mathbf{w}}_{u_i}^p$ into a matrix and obtain the positive intent kernel $\mathbf{w}_{u_i}^p \in \mathbb{R}^{d\times d}$:

$$\mathbf{w}_{u_i}^p = reshape(\tilde{\mathbf{w}}_{u_i}^p). \tag{6}$$

4.1.2. Source-drive neg-meta network

Similar to the source-drive pos-meta network, we propose a source-drive neg-meta network to generate the negative intent kernel by considering users' negative transferable characteristics. In particular, we take $\mathbf{n}_{u_i}$ as input, and generate a vector $\tilde{\mathbf{w}}_{u_i}^n \in \mathbb{R}^{d^2}$ with the meta network $g^n(\cdot)$ as follows:

$$\tilde{\mathbf{w}}_{u_i}^n = g^n(\mathbf{n}_{u_i};\phi_n), \tag{7}$$

where ϕ_n are learnable parameters. We reshape $\tilde{\mathbf{w}}_{u_i}^n$ into a matrix and obtain the negative intent kernel $\mathbf{w}_{u_i}^n \in \mathbb{R}^{d\times d}$:

$$\mathbf{w}_{u_i}^n = reshape(\tilde{\mathbf{w}}_{u_i}^n). \tag{8}$$

4.1.3. Personalized bridge functions

For each user u_i^s in the source domain, after obtaining the positive intent kernel $\mathbf{w}_{u_i}^p \in \mathbb{R}^{d\times d}$ and the negative intent kernel $\mathbf{w}_{u_i}^n \in \mathbb{R}^{d\times d}$, we also incorporate a mixed intent kernel $\mathbf{w}_{u_i} \in \mathbb{R}^{d\times d}$ which is used to capture the comprehensive transferable characteristics of the user. Formally, the mixed intent kernel is defined as follows:

$$\mathbf{w}_{u_i} = \mathbf{w}_{u_i}^p \odot \mathbf{w}_{u_i}^n, \tag{9}$$

where $\odot$ denotes the element-wise product. We leverage $\mathbf{w}_{u_i}^p$, $\mathbf{w}_{u_i}^n$ and $\mathbf{w}_{u_i}$ as the corresponding parameters of the positive personalized bridge function $f^{per}(\cdot;\mathbf{w}_{u_i}^p)$, the negative personalized bridge function $f^{per}(\cdot;\mathbf{w}_{u_i}^n)$ and the mixed personalized bridge function $f^{per}(\cdot;\mathbf{w}_{u_i})$, respectively. Following [8,22], we adopt a linear layer as $f^{per}(\cdot)$. With the three bridge functions, we can obtain the transformed representations in the target domain of $\mathbf{u}_i^s$, $\mathbf{p}_{u_i}$, and $\mathbf{n}_{u_i}$ as follows:

$$\mathbf{u}_i^{t_p} = f^{per}(\mathbf{u}_i^s;\mathbf{w}_{u_i}), \tag{10}$$

$$\mathbf{p}_{u_i}^{t_p} = f^{per}(\mathbf{p}_{u_i};\mathbf{w}_{u_i}^p), \tag{11}$$

$$\mathbf{n}_{u_i}^{t_p} = f^{per}(\mathbf{n}_{u_i};\mathbf{w}_{u_i}^n). \tag{12}$$

4.2. Common bridge

The above-mentioned tri-kernel personalized bridge is designed to model various relationships between user preferences in different domains, which can be considered as fine-grained user transferable characteristics. However, many users would share similar relationships in different domains where the common user transferable characteristics should be taken into consideration. To the end, we also introduce a common bridge which attempts to capture the coarse-grained user transferable characteristics.

More precisely, we take a two-layer feed-forward network $f^{com}(\cdot;\mathbf{w}_c)$ as the common bridge, where $\mathbf{w}_c$ are learnable parameters. For each user's characteristics in the source domain, i.e., $\mathbf{u}_i^s$, $\mathbf{p}_{u_i}$, and $\mathbf{n}_{u_i}$, we can obtain the corresponding transformed representations in the target domain as follows:

$$\mathbf{u}_i^{t_c} = f^{com}(\mathbf{u}_i^s;\mathbf{w}_c), \tag{13}$$

$$\mathbf{p}_{u_i}^{t_c} = f^{com}(\mathbf{p}_{u_i};\mathbf{w}_c), \tag{14}$$

$$\mathbf{n}_{u_i}^{t_c} = f^{com}(\mathbf{n}_{u_i};\mathbf{w}_c). \tag{15}$$

4.3. Target bridge

Since both tri-kernel personalized bridge and common bridge overlook the influence of the preference signals from the target domain, we argue that the target information could also bring useful signal to guide the process of preference transfer. To model the information from the target domain explicitly, we introduce a novel target bridge, which is developed to capture the complex relationships between the

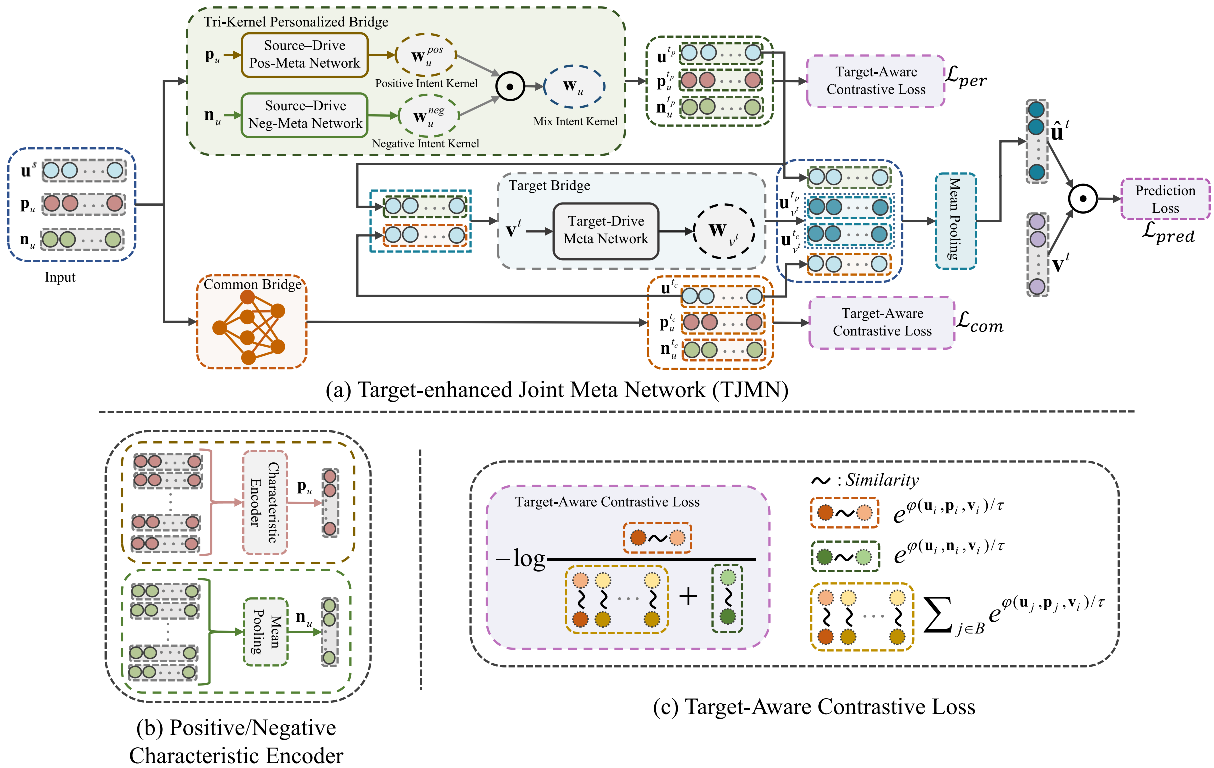

Fig. 2. The proposed framework TJMN for cross-domain recommendation. (a) The main procedure of TJMN, which attempts to learn multiple transfer bridges with different granularities of user preference as well as involve the information from the target domain to guide the learning process. (b) The structure of Positive/Negative Characteristic Encoder. (c) The structure of Target-Aware Contrastive Loss.

user preference of the source and target domain by capturing the characteristics of each target item in the target domain. Specifically, we leverage a meta network which takes a target item embedding $\mathbf{v}_j^t$ as the input, and the output $\tilde{\mathbf{w}}_{v_j^t} \in \mathbb{R}^{d^2}$ is reshaped into a matrix $\mathbf{w}_{v_j^t} \in \mathbb{R}^{d\times d}$. Formally, we have:

$$\tilde{\mathbf{w}}_{v_j^t} = g^t(\mathbf{v}_j^t; \phi_t), \tag{16}$$

$$\mathbf{w}_{v_j^t} = reshape(\tilde{\mathbf{w}}_{v_j^t}), \tag{17}$$

where ϕ_t are learnable parameters and $g^t(\cdot)$ is a two-layer feed-forward network. We utilize $\mathbf{w}_{v_j^t}$ as the corresponding parameters of the personalized target bridge function $f^{tar}(\cdot; \mathbf{w}_{v_j^t})$, and obtain the transformed target representations of $\mathbf{u}_i^{t_p}$ and $\mathbf{u}_i^{t_c}$ as follows:

$$\mathbf{u}_{i,v_j^t}^{t_p} = f^{tar}(\mathbf{u}_i^{t_p}; \mathbf{w}_{v_j^t}), \tag{18}$$

$$\mathbf{u}_{i,v_j^t}^{t_c} = f^{tar}(\mathbf{u}_i^{t_c}; \mathbf{w}_{v_j^t}). \tag{19}$$

4.4. Target-aware contrastive learning layer

In this section, we further incorporate contrastive [31] learning to enhance user representations in the target domain. Different from previous work, we introduce the information of the target item into the contrastive learning procedure and propose a target-aware contrastive learning layer. In particular, in each mini-batch $\mathcal{B}$, we have N samples with each sample $(u_i^o, S_{u_i}^p, S_{u_i}^n, v_j^t, r_{ij})$ consisting of a overlapping user $u_i^o \in U^o$, his corresponding positive items $S_{u_i}^p$ and negative interaction items $S_{u_i}^n$ in the source domain, a target item v_j^t and his rating score r_{ij} on v_j^t in the target domain. For the user u_i^o, we first obtain his transformed representations $\mathbf{u}_i^{t_p}$, $\mathbf{p}_{u_i}^{t_p}$, $\mathbf{n}_{u_i}^{t_p}$ via the tri-kernel personalized bridge. Then we leverage contrastive learning to learn better representations of u_i^o. Formally, we define the similarity between $\mathbf{u}_i^{t_p}$ and $\mathbf{p}_{u_i}^{t_p}$ as follows:

$$\phi(\mathbf{u}_i^{t_p}, \mathbf{p}_{u_i}^{t_p}, \mathbf{v}_j^t) = -(\hat{r}_{ij}^{uv} - \hat{r}_{ij}^{pv})^2, \tag{20}$$

where $\hat{r}_{ij}^{uv}$ is the inner-product of $\mathbf{u}_i^{t_p}$ and $\mathbf{v}_j^t$, and $\hat{r}_{ii}^{pv}$ is the inner product of $\mathbf{p}_{u_i}^{t_p}$ and $\mathbf{v}_j^t$. It is worth noting that, here we consider the triple $(\mathbf{u}_i^{t_p}, \mathbf{p}_{u_i}^{t_p}, \mathbf{v}_j^t)$ as a positive instance, where the target information $\mathbf{v}_j^t$ is introduced to assist the contrastive learning procedure. For the negative instance construction, we introduce two kinds of negative instances, i.e., $(\mathbf{u}_i^{t_p}, \mathbf{n}_{u_i}^{t_p}, \mathbf{v}_j^t)$ and $(\mathbf{u}_k^{t_p}, \mathbf{p}_{u_k}^{t_p}, \mathbf{v}_j^t)$, where the former represents negative instances based on the negative interaction items of u_i^o, and the latter considers the positive interaction items of other users in the same mini-batch as negative instances. Formally, the target-aware contrastive loss corresponding to the tri-kernel personalized bridge is defined as follows:

$$\mathcal{L}_{per} = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} \log \frac{e^{\phi(\mathbf{u}_i^{t_p}, \mathbf{p}_{u_i}^{t_p}, \mathbf{v}_j^t)/\tau}}{Z_{per}^1 + Z_{per}^2}, \tag{21}$$

where $Z_{per}^1 = e^{\phi(\mathbf{u}_i^{t_p}, \mathbf{n}_{u_i}^{t_p}, \mathbf{v}_j^t)/\tau}$, $Z_{per}^2 = \sum_{k\in B} e^{\phi(\mathbf{u}_k^{t_p}, \mathbf{p}_{u_k}^{t_p}, \mathbf{v}_j^t)/\tau}$, and τ denotes a temperature parameter.

In addition, we also incorporate a target-aware contrastive learning layer towards the transformed representations $\mathbf{u}_i^{t_c}$, $\mathbf{p}_{u_i}^{t_c}$, $\mathbf{n}_{u_i}^{t_c}$ of the common bridge. Similar to the $\mathcal{L}_{per}$, the target-aware contrastive loss corresponding to the common bridge is defined as follows:

$$\mathcal{L}_{com} = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} \log \frac{e^{\phi(\mathbf{u}_i^{t_c}, \mathbf{p}_{u_i}^{t_c}, \mathbf{v}_j^t)/\tau}}{Z_{com}^1 + Z_{com}^2}, \tag{22}$$

where $Z_{com}^1 = e^{\phi(\mathbf{u}_i^{t_c}, \mathbf{n}_{u_i}^{t_c}, \mathbf{v}_j^t)/\tau}$ and $Z_{com}^2 = \sum_{k\in B} e^{\phi(\mathbf{u}_k^{t_c}, \mathbf{p}_{u_k}^{t_c}, \mathbf{v}_j^t)/\tau}$.

4.5. Prediction layer

After the representations learned by different bridge functions, we obtain users' transformed representations in the target domain. Since a

Table 2
Statistics of 6 different CDR tasks. “Overlap” denotes the number of overlapping users.

Dataset		Amazon			Douban		
CDR tasks		Task1	Task2	Task3	Task4	Task5	Task6
domain	Source	Movie	Book	Book	DoubanBook	DoubanBook	DoubanMusic
	Target	Music	Movie	Music	DoubanMovie	DoubanMusic	DoubanMovie
Item	Source	50,052	367,982	367,982	95,872	95,872	79,878
	Target	64,443	50,052	64,443	34,893	79,878	34,893
User	Overlap	18,031	37,388	16,738	2,209	1,736	1,815
	Source	123,960	603,668	603,668	2,212	2,212	1,820
	Target	75,258	123,960	75,258	2,712	1,820	2,712
Rating	Source	1,697,533	8,898,041	8,898,041	227,251	227,251	179,847
	Target	1,097,592	1,697,533	1,097,592	1,278,401	179,847	1,278,401

user's final decision on candidate items heavily relies on his preference in the target domain, we fuse the transformed representations from different perspectives, including $\mathbf{u}_i^{t_p}$, $\mathbf{u}_i^{t_c}$, $\mathbf{u}_{i,v_j^t}^{t_p}$, $\mathbf{u}_{i,v_j^t}^{t_c}$. We adopt a mean-pooling layer to fuse these transformed representations, and the final transformed representation $\hat{\mathbf{u}}_i^t$ is represented as follows:

$$\hat{\mathbf{u}}_i^t = \frac{1}{4}(\mathbf{u}_i^{t_p} + \mathbf{u}_i^{t_c} + \mathbf{u}_{i,v_j^t}^{t_p} + \mathbf{u}_{i,v_j^t}^{t_c}). \tag{23}$$

Then, the predicted rating score $\hat{r}_{ij}$ is calculated by the inner-product of $\hat{\mathbf{u}}_i^t$ and the target item embedding $\mathbf{v}_j^t$:

$$\hat{r}_{ij} = (\hat{\mathbf{u}}_i^t)^{\mathrm{T}}\mathbf{v}_j^t. \tag{24}$$

Let $\mathcal{R}_o^t = \{r_{ij} | u_i \in U^o \wedge v_j \in V^t\}$ denote the ratings of overlapping users in the target domain, the loss function is formulated as:

$$\mathcal{L}_{pred} = \frac{1}{|\mathcal{R}_o^t|}\sum_{r_{ij}\in\mathcal{R}_o^t}(r_{ij} - \hat{r}_{ij})^2. \tag{25}$$

4.6. Model training

The proposed model is jointly trained by combining the task-oriented loss $\mathcal{L}_{pred}$ and the two target-aware contrastive losses $\mathcal{L}_{per}$ and $\mathcal{L}_{com}$. The overall loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{pred} + \lambda(\mathcal{L}_{per} + \mathcal{L}_{com}), \tag{26}$$

where λ is a trade-off parameter. We train all learnable parameters by minimizing $\mathcal{L}$ with the gradient descent method. It is worth noting that the model training is based on the overlapping users across the two domains. For both positive and negative characteristic encoders, we set the numbers of positive and negative items as 20. Following [2], we set the dimension of both users and items as 10. During the training process, all parameters are randomly initialized and optimized through back-propagation with the Adam algorithm. The learning rate is initialized as 0.02, and the mini-batch size for the Task 1, Task 2 and Task 3 are 128, 512 and 512, respectively. For all tasks, we set the trade-off parameter λ and the temperature parameter τ as 1 and 0.01, respectively. More analysis about the parameter settings will be discussed in Section 5.7.

5. Experiments

5.1. Datasets

We conduct our experiments on two real-world public datasets, i.e., the Amazon review dataset [2,14,22] and the Douban dataset [32,33]. For the Amazon review dataset, we choose three popular subsets, i.e., movies_and_tv (Movie), cds_and_viny1 (Music), and books (Book) for the experiments. Following [2], we define three CDR tasks, including Task 1: Movie→ Music, Task 2: Book→ Movie, and Task 3: Book→ Music. For the Douban dataset, we choose three subsets, i.e., DoubanBook, DoubanMusic, and DoubanMovie, and construct the other three CDR tasks, i.e., Task 4: DoubanBook→ DoubanMovie, Task 5: DoubanBook→ DoubanMusic, and Task 6: DoubanMusic→ DoubanMovie. All the 6 tasks will be used to investigate the effectiveness of the proposed approach.

For each task, we randomly sample a fraction of overlapping users as the test users and take the rest of overlapping users as the training set. Table 2 shows the statistics of the all tasks. In the experiment, we investigate three different proportions of the test users. To be specific, let β denote the proportion of the overlapping test users in all the overlapping users, we set β as 20%, 50%, and 80%, respectively. Note that a larger β indicates a severer data sparsity and cold-start problem.

5.2. Baselines

To evaluate the performance of the proposed approach, we compare it with the following six competitive baseline methods:

- **TGT** [34]. This baseline represents the target matrix factorization model and it is trained by solely utilizing the target domain data.
- **CMF** [5]. The collective matrix factorization (CMF) is an extension of matrix factorization model. It can simultaneously factor several matrices and share the user embeddings across both source and target domains.
- **EMCDR** [8]. It addresses the CDR problem by proposing an embedding-and-mapping framework. EMCDR first projects users and items in both source and target domains into two different latent spaces. After that, a mapping function is learned between the two latent spaces in order to model the coordinate relationship between the two domains.
- **DCDCSR** [9]. DCDCSR first employs the MF model to learn user and item latent factors, and then maps the latent factors across different domains using a fully connected deep neural network (DNN). To guide the training process of DNN, the rating sparsity degrees of individual users and items are also modeled for mapping latent factors across domains more accurately.
- **SSCDR** [22]. This baseline proposes a semi-supervised mapping based CDR framework, which can effectively learn the complex relationships across domains even in the case only a few number of labeled data is available. It first models the users and items in the metric spaces and then trains a cross-domain mapping function based on the distance-based loss defined by both unlabeled and labeled data. For these cold-start users, they infer their latent vectors by introducing a multi-hop neighborhood inference method.
- **PTUPCDR** [2]. This baseline proposes to employ the meta network to achieve personalized transfer of preference for users based on their characteristic embeddings. With the meta network, it generates personalized bridge function, which transforms user

Table 3
Effectiveness comparison in terms of the metric MAE (↓) between of our proposed approach TJMN and the state-of-the-art approaches. The best results are highlighted in bold, ∗ denotes the performance improvement over the best performing baseline (i.e., MIMNet) is statistically significant (t-test, p ≤ 0.05), and Imp% represents the relative improvement over MIMNet.

	β	TGT	CMF	DCDCSR	SSCDR	EMCDR	PTUPCDR	REMIT	MIMNet	TJMN	Imp%
Task1	20%	4.4803	1.5209	1.4918	1.3017	1.2350	1.1504	0.9393	0.7884	**0.7646**	3.02%
	50%	4.4989	1.6893	1.8144	1.3762	1.3277	1.2804	1.0437	0.8629	**0.7750***	10.19%
	80%	4.5020	2.4186	2.7194	1.5046	1.5008	1.4049	1.2181	1.0660	**0.8138***	23.66%
Task2	20%	4.1831	1.3632	1.3971	1.2390	1.1162	0.9970	0.8759	**0.8678**	0.8778	−1.15%
	50%	4.2288	1.5813	1.6731	1.2137	1.1832	1.0894	0.9172	0.8994	**0.8674***	3.56%
	80%	4.2123	2.1577	2.3618	1.3172	1.3156	1.1999	1.0055	0.9757	**0.8994***	7.82%
Task3	20%	4.4873	1.8284	1.8411	1.5414	1.3524	1.2286	1.3749	0.8221	**0.7448***	9.40%
	50%	4.5073	2.1282	2.1736	1.4739	1.4732	1.3764	1.4401	0.9271	**0.7578***	18.26%
	80%	4.6204	3.0130	3.1405	1.6414	1.7191	1.5784	1.6396	1.0782	**0.7835***	27.33%
Task4	20%	3.9377	1.1367	–	–	1.1492	0.8283	–	0.7138	**0.7135**	0.04%
	50%	4.0054	1.2362	–	–	1.0947	0.8590	–	**0.7300**	0.7340	−0.55%
	80%	3.9899	1.6421	–	–	1.2004	0.9478	–	0.7925	**0.7739***	2.35%
Task5	20%	4.4383	2.9467	–	–	3.1571	2.8758	–	1.9353	**1.1519***	40.48%
	50%	4.4037	3.2380	–	–	3.5488	3.3463	–	2.3345	**1.2314***	47.25%
	80%	4.3654	3.7770	–	–	3.7713	3.8538	–	2.6962	**1.3905***	48.43%
Task6	20%	4.0183	1.3738	–	–	1.4476	0.8357	–	**0.6945**	0.7123	−2.57%
	50%	4.0026	1.4496	–	–	1.4557	0.9112	–	0.7282	**0.7215***	0.92%
	80%	4.0056	2.0336	–	–	1.5702	1.0100	–	0.7735	**0.7670**	0.84%

preference embeddings in the source domain into the target domain. For the cold-start users in the target domain, PTUPCDR utilizes the transformed embeddings as their initial embeddings.

- **REMIT** [23]. REMIT considers users' multiple interests in the source domain. Specifically, it employs heterogeneous information network and meta-path to get users' multiple interests in the source domain and develops a reinforced method to dynamically assign weights to transformed multiple interests for different training instances to optimize the performance of the target model.
- **MIMNet** [24]. This is the state-of-the-art baseline, which also take into account the multiple interests of users. Specifically, it employs capsule network to learn user multiple interests and then transfers user representations from source domain to target domain based on the multiple interest-level preference bridge generated by the learned interests. As for the transferred user representations, MIMNet incorporates a multi-granularity target-guided attention network to aggregate them for recommendation.

5.3. Evaluation metrics

Following [2,8,35], we use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as the evaluation metrics. It is worth noting that a lower value of MAE and RMSE indicates a better model performance. Let n denote the number of samples, $\hat{y}_i$ and y_i denote the predicted and observed rating scores of the ith sample, respectively. For each metric, we repeat all experiments five times and report the average results to keep reliability. The definition of the two metrics are given as follows:

- **MAE** (Mean Absolute Error): It measures the average of the absolute error between the predicted and observed rating scores.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|. \tag{27}$$

- **RMSE** (Root Mean Square Error): It measures the sample standard deviation of the difference between the predicted and observed rating scores.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}. \tag{28}$$

5.4. Overall performance

In this section, we investigate the performance of our proposed approach TJMN under different cold-start levels. The overall performance of our proposed approach TJMN and the baselines on six tasks are illustrated in Tables 3 and 4 in terms of both metrics. The best results in terms of the corresponding metric are highlighted in bold, and ∗ denotes the performance improvement over the best performing baseline (i.e., MIMNet) is statistically significant (paired t-test, p ≤ 0.05). The results show that our approach achieves the best performance on most tasks. In addition, we have the following insights and analysis.

With a higher cold-start level (i.e., a larger value of β), the performance of all methods will decrease generally. This is because the task will become more challenging when less useful information is available. It is also interesting to observe that the relative performance improvements of our model over the best performing baseline (i.e., MIMNet) become larger with the increment of the β value in most cases. For example, on the task 1, the relative performance improvement in terms of the metric MAE of TJMN over MIMNet is 3.02% when $\beta = 20\%$, while the relative performance improvement reaches to 23.66% when $\beta = 80\%$. The results demonstrate that our model is more robust when training data becomes more sparse.

In addition, we can observe that TGT shows the worst performance. The reason is that TGT solely relies on the target domain and overlooks the rich information in the source domain. Compared to TGT, CMF obtains a better performance since it simultaneously models the information from both source and target domains by combining data from different domains into a single domain. This also indicates that exploiting the data from source domain is effective to alleviate the data sparsity problem and improve the recommendation performance in the target domain. However, the main limitation of CMF is that it ignores the difference of user preference in distinct domains and inevitably leads to inferior performance. The methods (i.e., EMCDR, DCDCSR, SSCDR, PTUPCDR), which attempt to learn a cross-domain mapping function, demonstrates a superior performance to CMF. The reason is attributed to that they can effectively alleviate the domain drift issue by modeling the complex relationships across domains. Among all baselines, MIMNet presents the best performance. This is because it bridges the source and target domains in a multi-interest and multi-granularity way, i.e., it employs the capsule network and multiple interest-level preference bridges to learn and transfer users' multiple interests, and

Table 4
Effectiveness comparison in terms of the metric RMSE (↓) between of our proposed approach TJMN and the state-of-the-art approaches. The best results are highlighted in bold, ∗ denotes the performance improvement over the best performing baseline (i.e., MIMNet) is statistically significant (t-test, p ≤ 0.05). lower is better, and Imp% represents the relative improvement over MIMNet.

	β	TGT	CMF	DCDCSR	SSCDR	EMCDR	PTUPCDR	REMIT	MIMNet	TJMN	Imp%
Task1	20%	5.1580	2.0158	1.9210	1.6579	1.5515	1.5195	1.2709	1.1020	**1.0426***	5.39%
	50%	5.1736	2.2271	2.3439	1.7477	1.6644	1.6380	1.4580	1.2185	**1.0600***	13.01%
	80%	5.1891	5.1891	3.3065	1.9229	1.8771	1.8234	1.6601	1.4830	**1.1236***	24.23%
Task2	20%	4.7536	1.7918	1.7346	1.6526	1.4120	1.3317	1.1650	1.1431	**1.1342**	0.78%
	50%	4.7920	2.0886	2.0551	1.5602	1.4981	1.4395	1.2379	1.1869	**1.1450***	3.53%
	80%	4.8149	2.6777	2.7702	1.7024	1.6433	1.5916	1.3772	1.3078	**1.1735***	10.27%
Task3	20%	5.1672	1.3829	2.2955	1.9283	1.6737	1.6085	1.9940	1.1487	**1.0137***	11.75%
	50%	5.1727	2.7275	2.6771	1.8441	1.8000	1.7447	2.0495	1.2924	**1.0297***	20.33%
	80%	5.2308	3.6948	3.5842	2.1403	2.1119	2.0510	2.2653	1.5029	**1.0736***	28.56%
Task4	20%	4.6492	1.5761	–	–	1.6556	1.1406	–	0.9356	**0.9225**	1.40%
	50%	4.7380	1.7325	–	–	1.5753	1.2376	–	0.9788	**0.9739**	0.50%
	80%	4.7240	2.2018	–	–	1.6889	1.3828	–	1.1225	**1.0353***	7.77%
Task5	20%	4.6428	3.7837	–	–	3.5161	3.7392	–	2.5115	**1.4966***	40.41%
	50%	5.1867	4.0923	–	–	3.8002	4.1313	–	2.9203	**1.5713***	46.19%
	80%	5.2015	4.5475	–	–	4.1031	4.4201	–	3.2734	**1.7159***	47.58%
Task6	20%	4.7354	1.9329	–	–	2.0317	1.1400	–	**0.8954**	0.9238	−3.17%
	50%	4.7355	2.0262	–	–	2.0353	1.3530	–	0.9766	**0.9409***	3.66%
	80%	4.7500	2.6492	–	–	2.1326	1.4437	–	1.0640	**1.0213***	4.01%

Table 5
Ablation study of our proposed TJMN in terms of the metric MAE (↓). The best results are highlighted in bold.

		w/o Con-Loss	w/o Target Bridge	w/o Common Bridge	w/o TKPB	Full
Task1	20%	0.7732	1.0779	0.9745	0.7981	**0.7646**
	50%	0.7809	1.1049	0.9757	0.8083	**0.7750**
	80%	0.8283	1.2421	1.0214	0.8378	**0.8138**
Task2	20%	0.8846	0.9800	1.1569	0.8845	**0.8778**
	50%	0.8842	0.9786	1.1916	0.8828	**0.8674**
	80%	0.9098	1.0422	1.2120	0.9078	**0.8994**
Task3	20%	0.7585	0.9441	1.1721	0.7610	**0.7448**
	50%	0.7791	1.0224	1.1806	0.7699	**0.7578**
	80%	0.8049	1.1057	1.2473	0.7871	**0.7835**

then utilize a novel multi-granularity target-guided attention module to aggregate transformed interest-level representations.

Our method TJMN significantly outperforms all baselines on all tasks generally in most cases. Specifically, when only 20% of the overlapping users are utilized for model training (i.e., $\beta = 80\%$), the relative performance improvements of TJMN over the best performing baseline MIMNet in terms of MAE and RMSE on the task 1 reach to 23.66% and 24.23%, respectively. Moreover, we observe that these state-of-the-art baselines such as PTUPCDR and MIMNet do not perform well on the task 5 compared to the other two tasks (task 4 and task 6) derived from the Douban dataset.For example, the MAE scores of PTUPCDR and MIMNet are 2.8758 and 1.9353 respectively when $\beta = 20\%$ on the task 5, while its corresponding MAE scores on the task 4 (task 6) are only 0.8283(0.8357) and 0.7138(0.6954) for PTUPCDR and MIMNet. In contrast, our proposed method presents a superior performance on task 5. Specifically, the relative performance improvements of TJMN over MIMNet on task 5 are 40.48%, 47.25%, 48.43% respectively when β equals to 20%, 50%, 80% in terms of MAE. This major reason is that our approach can effectively capture user preference by learning multiple transfer bridges with different granularities, as well as leverage the target information to guide the learning process.

5.5. Ablation study

To investigate the role of each component in our proposed TJMN, we perform an ablation study by removing each one from the entire model for comparison. The details of each variant are discussed as follows:

- **w/o Con-Loss**: The two target-aware contrastive losses, i.e., $\mathcal{L}_{per}$ and $\mathcal{L}_{com}$, will be removed from TJMN. In this case, only the prediction loss $\mathcal{L}_{pred}$ is utilized for guiding the training process.
- **w/o Target Bridge**: The target bridge will be removed from TJMN, which means that we do not involve user preference signals from the target domain to guide the process of preference transfer across different domains.
- **w/o Common Bridge**: We remove the common bridge from TJMN, which does not model the common user transferable characteristics between the source and target domains.
- **w/o TKPB**: We discard the whole tri-kernel personalized bridge module. In this case, the preference signals of different users will be overlooked in the preference transfer learning process.
- **Full**: It is the full model, i.e., TJMN, proposed in this paper.

The results on three Amazon review tasks are shown in Table 5 and Table 6. From the results we can observe that our proposed model TJMN (i.e., Full) achieves the best performance compared with any of its variants. To be specific, we have following observations: First, removing the two target-aware contrastive losses (i.e., w/o Con-Loss), the performance will degrade. For example, the performance will drop from 0.7646 (1.0426) to 0.7732 (1.0472) on task 1 when $\beta = 20\%$ in terms of the metric MAE (RMSE). Note that a higher value of MAE or RMSE corresponds a worse model performance. The results verify the validity of introducing the two contrastive losses. Second, discarding the target bridge (i.e., w/o Target Bridge) shows a considerable decline of performance, which proves the effectiveness of introducing the user preference information from the target domain to guide the preference transfer learning process. Third, utilizing the common bridge will enhance the performance of our approach as it captures the coarse-grained user transferable characteristics. At last, incorporating the tri-kernel personalized bridge can further boost the model performance because of the ability to model the fine-grained user transferable characteristics.

Table 6
Ablation study of our proposed TJMN in terms of the metric RMSE (↓). The best results are highlighted in bold.

		w/o Con-Loss	w/o Target Bridge	w/o Common Bridge	w/o TKPB	Full
Task1	20%	1.0472	1.3940	1.2726	1.0673	**1.0426**
	50%	1.0876	1.4407	1.2809	1.0831	**1.0600**
	80%	1.1388	1.6492	1.3495	1.1543	**1.1236**
Task2	20%	1.1376	1.2590	1.4717	1.1392	**1.1342**
	50%	1.1548	1.2612	1.5075	1.1493	**1.1450**
	80%	1.1956	1.3558	1.5374	1.1771	**1.1735**
Task3	20%	1.0297	1.2336	1.5016	1.0202	**1.0137**
	50%	1.0475	1.3418	1.5215	1.0347	**1.0297**
	80%	1.1025	1.4862	1.6108	1.0769	**1.0736**

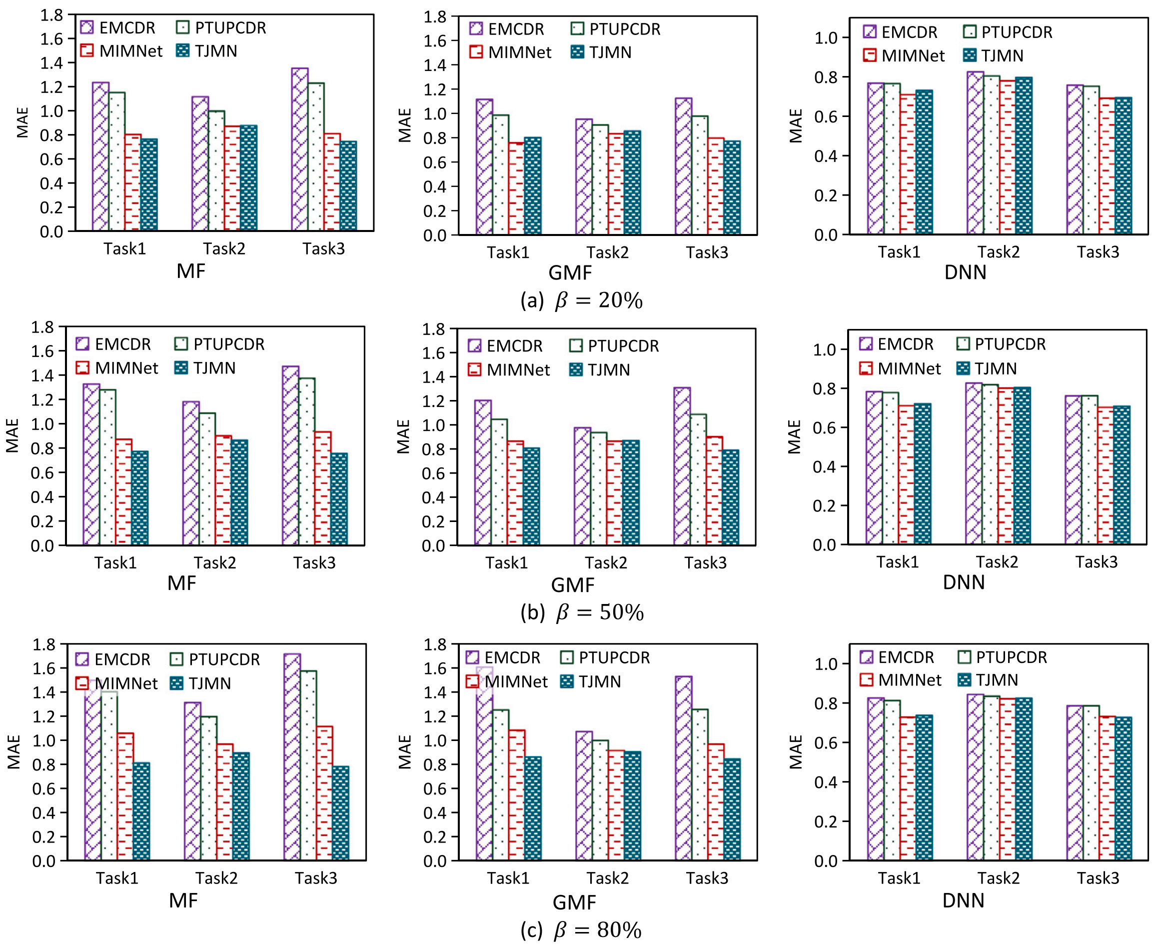

Fig. 3. Performance of the three state-of-the-art baselines EMCDR, PTUPCDR, MIMNet and our proposed model TJMN upon three different base models, including MF, GMF and YouTube DNN, with different β values (i.e., 20%, 50% and 80%) from top to bottom.

5.6. Generalization study

Existing CDR methods, including the proposed method TJMN, apply MF as the base model to conduct experimental evaluation. In this section, we investigate the performance of TJMN under different base models. Following [2], we testify the compatibility of TJMN by further utilizing other two base models, i.e., GMF [36] and YouTube DNN [37].

Fig. 3 demonstrates the performance of our proposed model TJMN together with the three state-of-the-art baselines EMCDR, PTUPCDR and MIMNet upon three different base models, including MF, GMF and YouTube DNN, with different β values (i.e., 20%, 50% and 80%). From the results, we have the following insightful observations: First, TJMN is consistently superior to both EMCDR and PTUPCDR on all tasks, and superior to MIMNet in most cases when different base models are utilized. This indicates that our proposed model has a good compatibility under different setting of base models. Second, TJMN upon the neural base models (i.e., GMF and YouTube DNN) shows considerably better performance as compared with the non-neural base model (i.e., MF). This is rationale as the neural base models are more powerful in large-scale real-world recommendations [2]. Third, the impact of data sparsity towards TJMN upon different base models is less sensitive as compared to both EMCDR and PTUPCDR. For example, when the data becomes more sparse (e.g., a higher value of β), the performance degradation of both EMCDR and PTUPCDR is significantly larger than that of TJMN.

5.7. Parameter sensitivity

Here, we investigate the impact of important hyperparameter settings on the performance of TJMN, including the trade-off parameter

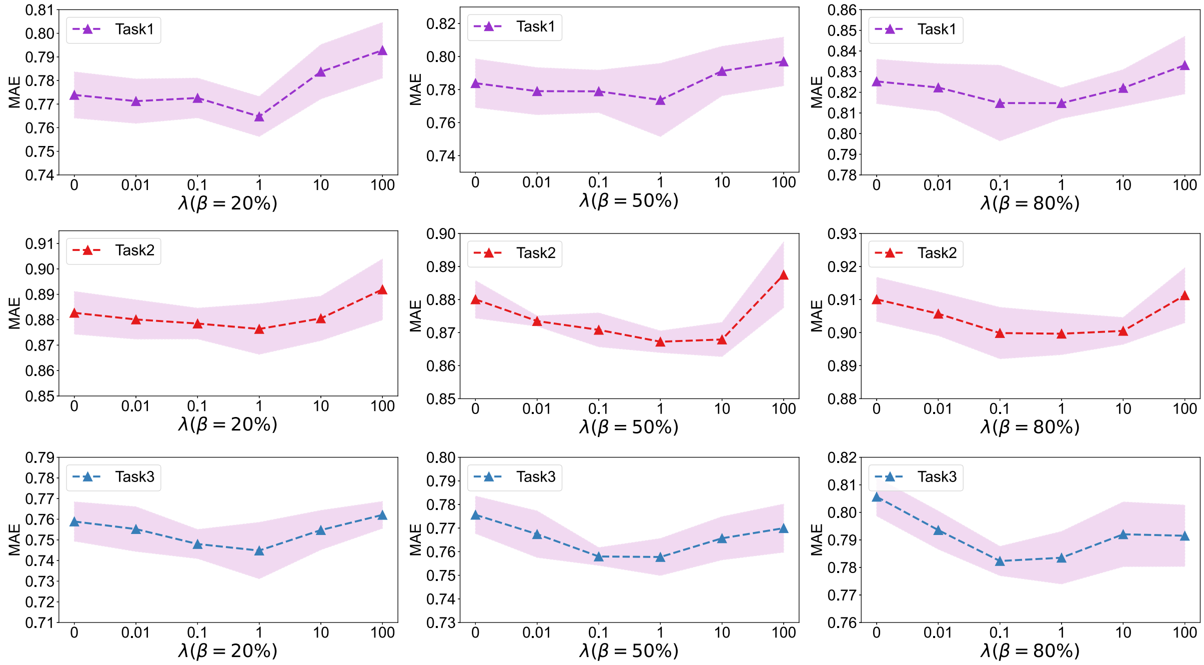

Fig. 4. Impact of λ on three Amazon review tasks in terms of MAE.

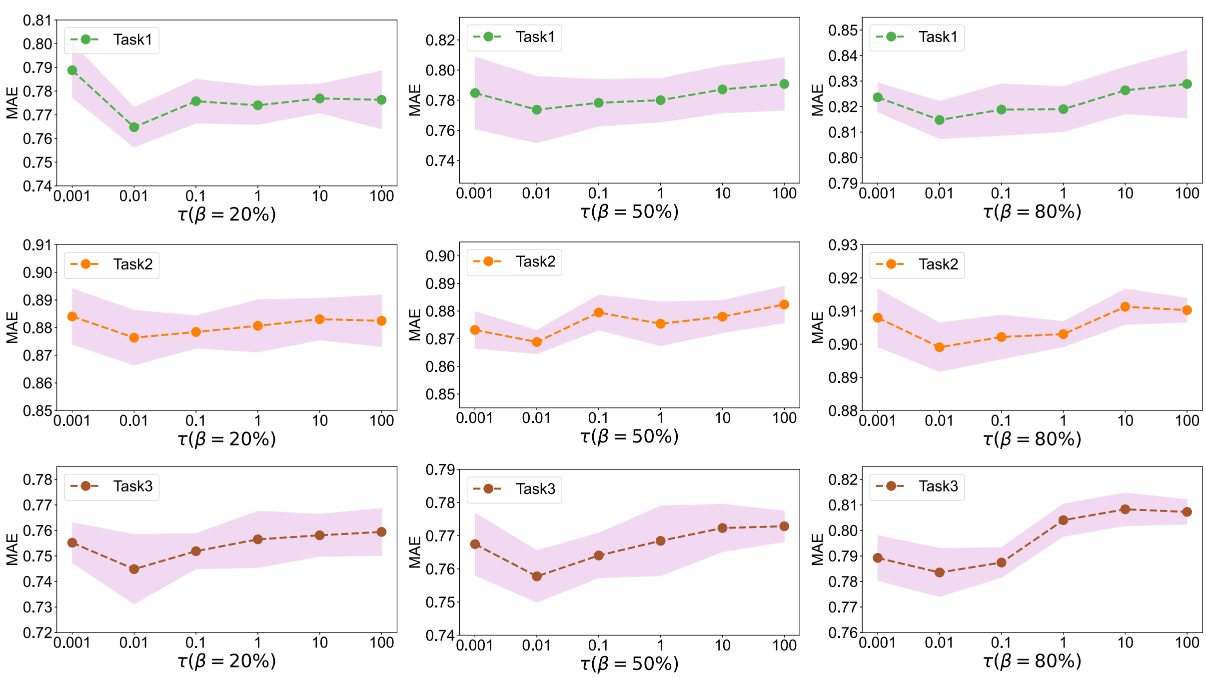

Fig. 5. Impact of τ on three Amazon review tasks in terms of MAE.

λ and temperature parameter τ. The results on three Amazon review tasks in terms of MAE are shown in Fig. 4 and Fig. 5, respectively.

- **Impact of the parameter** λ. The trade-off parameter λ in Eq. (26) controls the importance of the target-aware contrastive loss. We vary λ in {0, 0.01, 0.1, 1, 10, 100} and the results are shown in Fig. 4. On Task 1, we can see that the MAE performance of TJMN first increases gradually when we raise the value of λ, and reaches a peak when $\lambda = 1$. If we continue to increase λ, the model performance will drop considerably. Similar trends can also be observed on other two tasks. The results demonstrate that incorporating the two target-aware contrastive loss is critical for learning better transformed representations, while focusing too much on this loss will also deteriorate the performance of our proposed model.
- **Impact of temperature parameter** τ. The temperature parameter τ controls the concentration level of representations. We vary τ in {0.001, 0.01, 0.1, 1, 10, 100} by fixing $\lambda = 1$, and the results are reported in Fig. 5. On all three tasks, we can observe that the performance of our model first raises slightly when we gradually increase the value of τ and reaches a peak when a relatively small

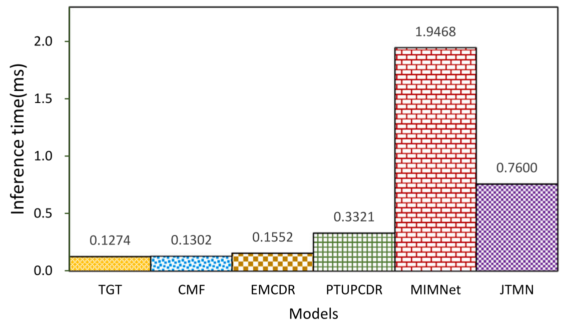

Fig. 6. Inference time (milliseconds) for a batch of samples with a batch size of 128 for different methods on Task 1.

value of τ is used, e.g., $\tau = 0.01$. If we continue to increase the value of τ, there will be a considerable degradation of model performance. The reason is that a small τ value corresponds to a higher gradient for model training with respect to these hard negative interaction items, and the penalty will concentrate more on the high similarity region.

5.8. Efficiency analysis

In this section, we investigate the efficiency of our proposed method JTMN. Fig. 6 shows the inference time of different methods. We can observe that the three comparing methods, i.e., TGT, CMF and EMCDR, are most efficient as they mainly rely on the matrix factorization techniques. PTUPCDR demonstrates a higher inference time compared to them. This is because PTUPCDR needs to generate a personalized bridge for each user. The inference time of MIMNet is the highest among all comparing methods as it attempts to apply multiple interest bridges for each user. In addition, it also needs to extract the target prototype-level signal for each target item to guide the adaptive aggregation process of user preference in the target domain. From Fig. 6, we can see that the inference time of JTMN is clearly lower than the best performing baseline MIMNet, and slightly higher than the remaining methods. The result shows that our proposed method is efficient, making it practicable in potential applications.

6. Conclusion

In this paper, we propose the target-enhanced joint meta network with contrastive learning for solving the cold-start problem in cross-domain recommendation. To effectively guide the user preference transfer across different domains, we inject the signals from the target domain to the transferring process and propose a novel target bridge. Moreover, we simultaneously utilize multi-grained transfer bridges to capture the complex relationships of user preference between the source and target domains. In addition, we model the rich partial order relation information via exploiting these negative interacted items and propose a tri-kernel personalized bridge. Finally, we also propose a target-aware contrastive learning layer to learn more informative representations towards the target domain. Experiments on six CDR tasks verified the effectiveness of our proposed approach. The ablation results and generalization study further testify the contribution of each component and the strong model compatibility with different base models.

In future work, we will consider exploring user profiling to better catch their characteristics for preference transfer. In addition, another future work is to incorporate auxiliary information of items, such as attributes or multi-modality contents, to guide the representation learning of items.

CRediT authorship contribution statement

Xiaofei Zhu: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Lele Duan:** Writing – original draft, Software, Methodology, Conceptualization. **Stefan Dietze:** Writing – review & editing, Supervision. **Ran Yu:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant number 62472059]; the Natural Science Foundation of Chongqing, China [grant number CSTB2022NSCQ-MSX1672]; the Chongqing Talent Plan Project, China [grant number CSTC2024YCJH-BGZXM0022]; the Major Project of Science and Technology Research Program of Chongqing Education Commission of China [grant number KJZD-M202201102].

Data availability

Data will be made available on request.

References

[1] G. Tang, X. Zhu, J. Guo, S. Dietze, Time enhanced graph neural networks for session-based recommendation, Knowl.-Based Syst. 251 (2022) 109204.
[2] Y. Zhu, Z. Tang, Y. Liu, F. Zhuang, R. Xie, X. Zhang, L. Lin, Q. He, Personalized transfer of user preferences for cross-domain recommendation, in: Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022, pp. 1507–1515.
[3] R. Guan, H. Pang, F. Giunchiglia, Y. Liang, X. Feng, Cross-domain meta-learner for cold-start recommendation, IEEE Trans. Knowl. Data Eng. 35 (2023) 7829–7843.
[4] W. Yang, Y. Jian, Y. Wang, S. Lu, L. Shen, B. Wang, H. Tang, L. Zhang, Not all embeddings are created equal: Towards robust cross-domain recommendation via contrastive learning, in: Proceedings of the ACM on Web Conference 2024, 2024, pp. 3195–3206.
[5] A.P. Singh, G.J. Gordon, Relational learning via collective matrix factorization, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 650–658.
[6] Y. Li, Q. Wu, L. Hou, J. Li, Entity knowledge transfer-oriented dual-target cross-domain recommendations, Expert Syst. Appl. 195 (2022) 116591.
[7] W. Liu, X. Zheng, M. Hu, C. Chen, Collaborative filtering with attribution alignment for review-based non-overlapped cross domain recommendation, in: Proceedings of the ACM Web Conference, vol. 2022, 2022, pp. 1181–1190.
[8] T. Man, H. Shen, X. Jin, X. Cheng, Cross-domain recommendation: An embedding and mapping approach, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 2464–2470.
[9] F. Zhu, Y. Wang, C. Chen, G. Liu, M.A. Orgun, J. Wu, A deep framework for cross-domain and cross-system recommendations, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018, pp. 3711–3717.
[10] H. Wang, Y. Zuo, H. Li, J. Wu, Cross-domain recommendation with user personality, Knowl.-Based Syst. 213 (2021) 106664.
[11] M. Andrychowicz, M. Denil, S.G. Colmenarejo, M.W. Hoffman, D. Pfau, T. Schaul, N.de. Freitas, Learning to learn by gradient descent by gradient descent, in: Proceedings of the 30th Conference on Neural Information Processing Systems, 2016, pp. 3981–3989.
[12] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 1126–1135.
[13] J. Lin, H. Shao, Z. Min, J. Luo, Y. Xiao, S. Yan, J. Zhou, Cross-domain fault diagnosis of bearing using improved semi-supervised meta-learning towards interference of out-of-distribution samples, Knowl.-Based Syst. 252 (2022) 109493.
[14] Y. Zhu, K. Ge, F. Zhuang, R. Xie, D. Xi, X. Zhang, L. Lin, Q. He, Transfer-meta framework for cross-domain recommendation to cold-start users, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1813–1817.

[15] N. Heidari, P. Moradi, A. Koochari, An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems, Knowl.-Based Syst. 256 (2022) 109835.

[16] X. Li, J. Sheng, J. Cao, W. Zhang, Q. Li, T. Liu, CDRNP: cross-domain recommendation to cold-start users via neural process, in: Proceedings of the 17th ACM International Conference on Web Search and Data Mining, 2024, pp. 378–386.

[17] Y. Xu, E. Wang, Y. Yang, H. Xiong, GS-rs: A generative approach for alleviating cold start and filter bubbles in recommender systems, IEEE Trans. Knowl. Data Eng. 36 (2024) 668–681.

[18] B. Li, X. Zhu, R. Li, C. Zhang, X. Xue, X. Wu, Cross-domain collaborative filtering over time, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, 2011, pp. 2293–2298.

[19] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, C. Zhu, Personalized recommendation via cross-domain triadic factorization, in: Proceedings of the 22nd International World Wide Web Conference, 2013, pp. 595–606.

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

[21] B. Li, Q. Yang, X. Xue, Transfer learning for collaborative filtering via a rating-matrix generative model, in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 617–624.

[22] S. Kang, J. Hwang, D. Lee, H. Yu, Semi-supervised learning for cross-domain recommendation to cold-start users, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1563–1572.

[23] C. Sun, J. Gu, B. Hu, X. Dong, H. Li, L. Cheng, L. Mo, REMIT: reinforced multi-interest transfer for cross-domain recommendation, in: Proceedings of 37th AAAI Conference on Artificial Intelligence, 2023, pp. 9900–9908.

[24] X. Zhu, Y. Yin, L. Wang, Mimnet: Multi-interest meta network with multi-granularity target-guided attention for cross-domain recommendation, 2024, arXiv:2408.00038.

[25] R. Jenatton, N.L. Roux, A. Bordes, G. Obozinski, A latent factor model for highly multi-relational data, in: 26th Annual Conference on Neural Information Processing Systems, in: Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 3176–3184.

[26] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, T. Chua, Attentional factorization machines: Learning the weight of feature interactions via attention networks, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 3119–3125.

[27] W. Xu, H. He, M. Tan, Y. Li, J. Lang, D. Guo, Deep interest with hierarchical attention network for click-through rate prediction, 2020, CoRR abs/2005.12981.

[28] S. Liu, S. Xu, W. Yu, Z. Fu, Y. Zhang, A. Marian, Fedct: Federated collaborative transfer for recommendation, in: SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021a, pp. 716–725.

[29] W. Liu, J. Su, C. Chen, X. Zheng, Leveraging distribution alignment via stein path for cross-domain cold-start recommendation, in: Annual Conference on Neural Information Processing System, in: Advances in Neural Information Processing Systems, vol. 34, 2021b, pp. 19223–19234.

[30] T. Munkhdalai, H. Yu, Meta networks, in: Proceedings of machine learning research, Vol. 70, 2017, pp. 2554–2563.

[31] B. Liang, Q. Zhu, X. Li, M. Yang, L. Gui, Y. He, R. Xu, Jointcl: A joint contrastive learning framework for zero-shot stance detection, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, pp. 81–91.

[32] F. Zhu, C. Chen, Y. Wang, G. Liu, X. Zheng, DTCDR: A framework for dual-target cross-domain recommendation, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November (2019) 3-7, 2019, pp. 1533–1542.

[33] F. Zhu, Y. Wang, C. Chen, G. Liu, X. Zheng, A graphical and attentional framework for dual-target cross-domain recommendation, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, pp. 3001–3008.

[34] Y. Koren, R.M. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (2009) 30–37.

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

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

[37] P. Covington, J. Adams, E. Sargin, Deep neural networks for youtube recommendations, in: Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 191–198.