Meta-Learning for Multimodal Cross-Domain Recommendation Systems

Yi Liu[1], **Xiaoyu Xu**[2], **Changjin Yu**[3]

{36920241153240 AI}[1] {23320241154660 Information} [2] {36920241153274 AI}[3]

Abstract

The cold-start problem is a critical challenge in recommender systems, especially for new users who lack sufficient interaction data. Cross-domain recommendation (CDR) methods have emerged as a promising solution, leveraging auxiliary source domains to enhance recommendations for target domains. However, existing CDR methods primarily depend on overlapping users between domains, which are often sparse in real-world datasets, leading to poor generalization and limited scalability. Additionally, these methods typically focus on learning bridging functions based on user-item interactions, neglecting the rich multimodal information, such as textual data, that could further improve the model's performance. To address these issues, we propose a novel MultiModal Cross-Domain Recommendation framework based on Meta-learning (MetaMMCDR). Our framework leverages multimodal item information to effectively model user preferences and learn personalized mapping functions that transfer these preferences across domains. By integrating multimodal data and meta-learning, MetaMMCDR mitigates the reliance on overlapping users and improves the generalization and performance of cold-start recommendations in cross-domain settings.

Introduction

In recent years, e-commerce and video recommendation platforms have undergone rapid growth, making the ability to recommend items that users enjoy increasingly crucial. However, the cold-start problem remains a significant challenge, as new users often fail to receive satisfactory recommendations, which negatively impacts user retention on these platforms.

Cross-domain recommendation (CDR) (Singh and Gordon 2008) has emerged as a promising solution to address this issue by leveraging rich information from auxiliary (source) domains to enhance the performance of recommendation systems in target domains.

Many studies have demonstrated the effectiveness of CDR methods based on embedding mapping (Kang et al. 2019; Man et al. 2017; Zhao et al. 2020; Fu et al. 2019; Zhu et al. 2018). However, these methods often encounter a critical issue: heavy reliance on overlapping users between domains. In such approaches, the number of training samples is determined by the number of overlapping users, which is typically low in real-world datasets. So, the learned mapping function often suffers from poor generalization, making it challenging to accurately represent user preferences in the target domain. Furthermore, these methods generally focus on user-item interactions when constructing embeddings, overlooking the rich multimodal information, such as text data, which could significantly improve knowledge transfer. To address these challenges, we propose a novel MultiModal Cross-Domain Recommendation framework based on meta-learning (MetaMMCDR). The proposed framework consists of two key stages: a pre-training stage and a meta-learning stage.

During the pre-training stage, we leverage multimodal information to learn user embeddings separately for the source and target domains. Importantly, the data used in this stage is comprehensive and not limited to overlapping users, enabling the model to capture a broader range of user preferences.

The meta-learning stage aims to develop a meta network that takes user characteristic embeddings from the source domain as input and generates personalized mapping functions for each user. After training, we feed the source-domain user embeddings into these mapping functions to obtain transformed user embeddings.

Moreover, existing methods for learning mapping functions typically rely on a mapping-oriented optimization procedure that minimizes the distance between the embeddings of users from the source and target domains. However, such optimization is highly sensitive to the quality of the embeddings, which limits the performance of the learned mapping function. In contrast, our approach follows previous work (Tao et al. 2022) and employs a task-oriented optimization procedure that directly utilizes the rating task as the optimization goal, resulting in more robust and accurate mappings.

The main contributions of this work can be summarized in two key aspects:

- We introduce MetaMMCDR, a novel framework that integrates multimodal information and meta-learning to generate personalized mapping functions, addressing the cold-start problem in CDR.
- We plan to validate the effectiveness of MetaMMCDR

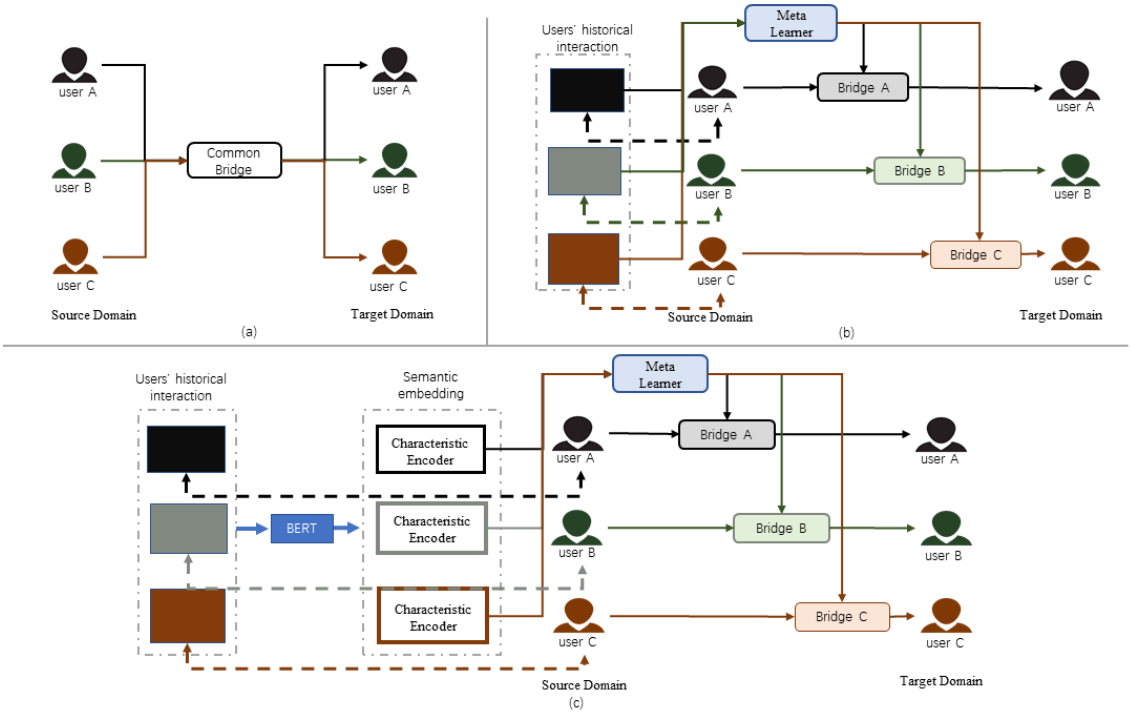

Figure 1: (a)In existing CDR methods: all users share the common bridge function. (b) PTUPCDR utilizes a meta network to generate personalized bridge functions for each user. (c) The proposed MetaMMCDR leverages multimodal item information to improve the generalization and performance.

through extensive experiments on cross-domain recommendation tasks using the Amazon dataset, aiming to demonstrate its ability to significantly improve recommendation performance.

Related Work

Cross-domain Recommendation

Cross-domain recommendation (CDR), inspired by transfer learning, has emerged as a promising solution to address the cold-start problem. In recent years, many deep learning-based models have been proposed to enhance knowledge transfer across domains (Hao et al. 2021; He et al. 2018; Hu, Zhang, and Yang 2018; Xi et al. 2021; Xie et al. 2022; Zhu et al. 2021c). Another important group of CDR methods focuses on bridging user preferences across different domains (Kang et al. 2019; Man et al. 2017; Pan et al. 2010; Zhao et al. 2020; Zhu et al. 2018; Lee et al. 2019; Zhu et al. 2021a), which is most closely related to our work. However, to the best of our knowledge, all existing methods learn mapping functions without incorporating multimodal information. In contrast, our proposed MetaMMCDR is the first approach to learn personalized bridges by integrating multimodal data.

Meta Learning for CDR

Meta-learning, also referred to as "learning to learn," aims to improve the performance of models on novel tasks by training them on a few-shot learning paradigm. Several meta-learning methods have been proposed to enhance the performance of recommender systems (Zhu et al. 2021c; Lee et al. 2019; Pan et al. 2019; Zhu et al. 2021b,a, 2022). Among these, the most relevant work includes TMCDR (Zhu et al. 2021a) and PTUPCDR (Zhu et al. 2022), both of which employ meta-learning techniques in CDR. However, similar to other CDR approaches, they do not incorporate multimodal information in the learning process.

Multimodal Recommendation

Multimodal recommendation systems are designed to understand and interpret data from various modalities, thereby addressing the issues of data sparsity and the cold-start problem. Current methods typically extract features from different modalities and fuse them to form a unified item representation (He and McAuley 2016; Wang et al. 2021; Tao et al. 2020, 2022). Given that users frequently engage with platforms containing rich multimodal information, it is essential to learn multimodal representations to improve the accuracy and relevance of recommendations.

Method

Problem Setting

In Cross-Domain Recommendation (CDR), we define two domains: a source domain and a target domain. Each domain consists of a set of users, $\mathcal{U} = \{u_1, u_2, \dots\}$, a set of

items, $\mathcal{V} = \{v_1, v_2, \dots\}$, and a rating matrix $\mathcal{R}$. The entry $r_{ij} \in \mathcal{R}$ represents the interaction between user u_i and item v_j. To distinguish between the two domains, we denote the user set, item set, and rating matrix of the source domain as $\mathcal{U}^s, \mathcal{V}^s, \mathcal{R}^s$, and the corresponding sets for the target domain as $\mathcal{U}^t, \mathcal{V}^t, \mathcal{R}^t$. The set of overlapping users is denoted as $\mathcal{U}^o = \mathcal{U}^s \cap \mathcal{U}^t$, while there is no overlap of items between the source and target domains.

Both users and items are represented as dense vectors, also known as embeddings or factors. In this work, the embeddings of a user u_i in domain $d \in \{s, t\}$ and item v_j are denoted as $u_i^d \in \mathbb{R}^k$ and $v_j^d \in \mathbb{R}^k$, where k is the dimensionality of the embeddings.

Semantic Embedding

Previous research has demonstrated that semantic embeddings of items are effective for knowledge transfer between domains, especially when user behaviors across domains are homogeneous or heterogeneous.

To obtain semantic item embeddings, we leverage pre-trained LLMs, such as BERT, to extract text representations that capture semantic information. Specifically, for each item v, we denote its textual content as $text(v)$ and its semantic embedding as $\mathbf{v}$. The semantic embedding $\mathbf{v}$ is computed as the sum of the token embeddings produced by BERT for each token in the item's description:

$$\mathbf{v} = \sum_{t \in text(v)} \phi_{BERT}(t), \tag{1}$$

where $\phi_{BERT}(t)$ represents the embedding of token t obtained from the BERT model.

Characteristic Encoder

The first step in generating the personalized bridge function is to extract users' transferable characteristics from the items they have interacted with. For cold-start users, who have no interaction history in the target domain, we rely on their interactions in the source domain. To capture the transferable characteristics of users, we apply an attention mechanism to the item embeddings, performing a weighted sum of item embeddings to allow different parts to contribute differently when aggregating them into a single representation. The characteristic embedding for user u_i is computed as:

$$p_{u_i} = \sum_{v_j^s \in \mathcal{S}_{u_i}} a_j v_j^s, \tag{2}$$

where $\mathcal{S}_{u_i}$ is the set of items that user u_i has interacted with in the source domain. The attention scores a_j are computed by applying a two-layer feed-forward network $h(\cdot)$:

$$a'_j = h(v_j; \theta), a_j = \frac{\exp(a'_j)}{\sum_{v_l^s \in \mathcal{S}_{u_i}} \exp(a'_l)}, \tag{3}$$

where a_j represents the attention score for item v_j, and θ denotes the parameters of the attention network $h(\cdot)$. The resulting user characteristic embedding $p_{u_i} \in \mathbb{R}^k$ represents the transferable features of user u_i based on their interactions with items in the source domain.

Meta Network

Users' preferences across domains vary significantly, and capturing these personalized preferences is crucial for effective cross-domain recommendation. To achieve this, we employ a meta network that takes the user's transferable characteristics, p_{u_i}, as input and generates personalized mapping functions to bridge the user embeddings in the source and target domains. The meta network is formulated as:

$$w_{u_i} = g(p_{u_i}; \phi), \tag{4}$$

where $g(\cdot)$ is the meta network parameterized by ϕ, and w_{u_i} is the personalized mapping vector for user u_i. This vector w_{u_i} is used as the parameter for the bridge function $f(\cdot)$. The bridge function can take various forms, and for simplicity, we use a linear layer for $f(\cdot)$, as in previous work (Man et al. 2017). The vector w_{u_i} is reshaped into a matrix $w_{u_i} \in \mathbb{R}^{k \times k}$, and with this bridge function, we obtain the transformed user embeddings in the target domain:

$$\hat{u}_i^t = f_{u_i}(u_t^s; w_{u_i}), \tag{5}$$

where $\hat{u}_i^t$ represents the transformed user embedding in the target domain, which can then be used for predictions. Notably, the vector w_{u_i} is used as the parameter of the bridge function, rather than as input. To train the meta network and characteristic encoder, we can minimize the distance using the mapping-oriented optimization procedure following existing bridge-based methods:

$$\mathcal{L} = \sum_{u_i \in \mathcal{U}^o} ||\hat{u}_i^t - u_i^t||^2, \tag{6}$$

and the task-oriented loss of our work can be formulated as:

$$\min_{\theta, \phi} \frac{1}{|\mathcal{R}_o^t|} \sum_{r_{ij} \in \mathcal{R}_o^t} (r_{ij} - f_{u_i}(u_i^s; w_{u_i})\mathbf{v}_j)^2, \tag{7}$$

where $\mathcal{R}_o^t = \{r_{ij} | u_i \in \mathcal{U}^o, v_j \in \mathcal{V}^t\}$ denotes the interactions of overlapping users in the target domain.

Algorithm 1: MultiModal Cross-Domain Recommendation framework based on Meta-learning (MetaMMCDR)

Input: $\mathcal{U}^s, \mathcal{U}^t, \mathcal{V}^s, \mathcal{V}^t, \mathcal{U}^o, \mathcal{R}^s, \mathcal{R}^t, \mathbf{V}^s, \mathbf{V}^t$.
Input: Meta network g_ϕ.
Input: Characteristic encoder h_θ.

Pre-training Stage:

1. Learning a source model which contains u^s, v^s.
2. Learning a target model which contains u^t, v^t.

Meta Stage:

3. Learning a characteristic encoder h_θ and a meta network g_ϕ by minimizing Equation 7.

Initialization Stage:

4. For a cold-start user u^t in the target domain, we use the transformed embedding $f_{u_i}(u_i^s; w_{u_i})$ as the user's initialized embedding in the target domain.

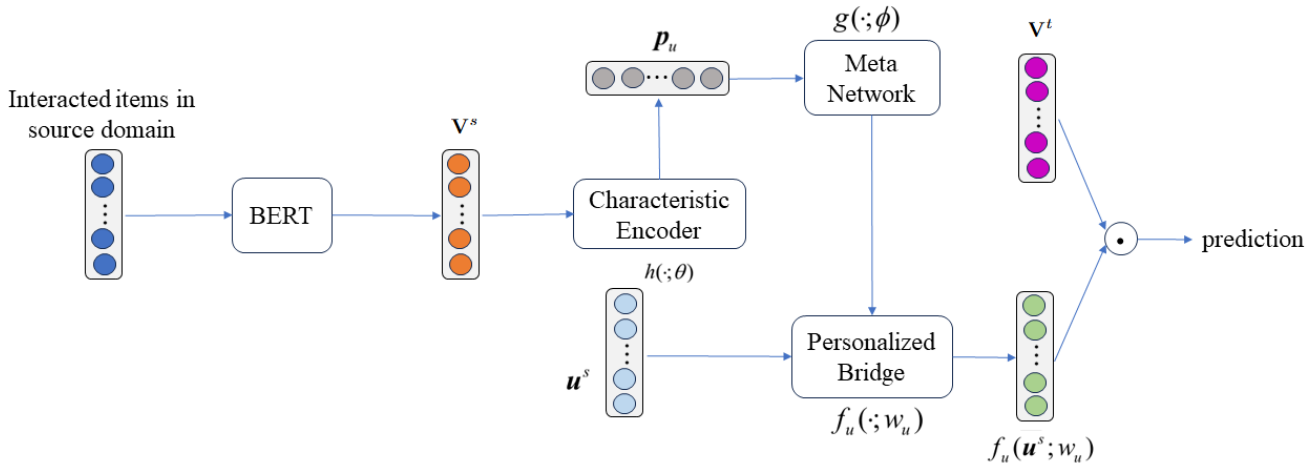

Figure 2: MetaMMCDR utilizes a meta network with users' characteristic embeddings in the source domain as input to generate personalized bridge functions for each user. Then, we can obtain the transformed user's embeddings as the initial embeddings in the target domain.

Overall Procedure

The structure of MetaMMCDR is shown in Figure 2. The training procedure can be divided into three steps: pretraining, meta and initialization stages, as see Algorithm 1.

Experiment

Experimental Settings

Dataset and Evaluation Metrics. Following most existing methods (Kang et al. 2019; Zhao et al. 2020; Zhu et al. 2021a, 2022), our experiment utilizes a real-world public dataset, namely the Amazon review dataset. Specifically, we use the Amazon-5cores dataset, where each user or item has at least five ratings. We choose three popular categories out of a total of 24: movies and TV (Movie), CDs and Vinyl (CD), and books (Book). We define three CDR tasks as follows: Task 1: Movie $\rightarrow$ CD, Task 2: Book $\rightarrow$ Movie, and Task 3: Book $\rightarrow$ CD. As shown in the details in Table 1, the number of ratings in the source domain is significantly larger than that in the target domain. In line with (Man et al. 2017; Zhao et al. 2020; Zhu et al. 2022), we adopt Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as the evaluation metrics for performance comparison.

Baseline. Since our method belongs to the bridge-based methods for CDR, we compare MetaMMCDR primarily with other bridge-based methods. Therefore, we select the following baseline methods for comparison:

- **Target**: Target denotes the target MF model, which is trained only using target domain data.
- **EMCDR** (Man et al. 2017): A popular CDR method for cold-start problems, which uses Matrix Factorization (MF) to learn embeddings and then employs a network to bridge the user embeddings from the source domain to the target domain.
- **SSCDR** (Kang et al. 2019): A semi-supervised, bridge-based CDR method.
- **PTUPCDR** (Zhu et al. 2022): This method trains a meta network using users' characteristic embeddings to learn personalized bridge functions, facilitating the personalized transfer of user preferences.

Implementation Details. Following (Man et al. 2017; Zhu et al. 2022), we randomly sample a fraction of ratings from overlapping users in the target domain and treat them as test users. We set the proportions of test(cold-start) users as 80%, while the remaining samples of overlapping users are used for training.

Cold-start Experiments

This section presents experimental results on cold-start scenarios. Table 2 presents the results of the experiments. The Target is a single-domain approach that solely relies on data from the target domain, which unfortunately leads to subpar performance. In contrast to TGT, various cross-domain

Table 1: Statistics of the cross-domain tasks

CDR Tasks	Domain		Item		User			Rating	
	Source	Target	Source	Target	Overlap	Source	Target	Source	Target
Task 1	Movie	CD	50,052	64,443	18,031	123,960	75,258	1,697,533	1,097,592
Task 2	Book	Movie	367,982	50,052	37,388	603,668	123,960	8,898,041	1,697,533
Task 3	Book	CD	367,982	64,443	16,738	603,668	75,258	8,898,041	1,097,592

Table 2: Cold-start results (MAE and RMSE) of 3 cross-domain tasks.

Source Domain	Target Domain	Metric	Target	SSCDR	EMCDR	PTUCDR	MetaMMCDR
Movie	CD	MAE	4.480	1.301	1.235	1.150	0.901
		RMSE	5.158	1.6579	1.551	1.519	1.172
Book	Movie	MAE	4.183	1.239	1.116	0.997	0.925
		RMSE	4.814	1.652	1.412	1.331	1.205
Book	CD	MAE	4.520	1.541	1.352	1.228	0.840
		RMSE	5.230	1.928	1.673	1.608	1.115

methodologies have the advantage of leveraging data from the source domain, thereby yielding superior outcomes. Consequently, harnessing data from an auxiliary domain emerges as an effective strategy to mitigate data scarcity and enhance recommendation accuracy within the target domain. Our method could significantly outperform the best baseline in most scenarios, demonstrating that MetaMMCDR is effective for cold start recommendation. The significant improvements achieved are due to the incorporation of additional information from the textual modality. It is worth noting that incorporating information from the visual modality simultaneously can lead to a decline in recommendation performance. This indicates that exploring how to properly introduce multimodal information is necessary for cross-domain recommendations.

Generalization Experiments

MF is a non-neural model, and it is probably too simple to achieve satisfying performance in large-scale real-world recommendations. Thus, to testify the compatibility of our framework as well as other bridge-based methods, we apply EMCDR, PTUPCDR and ours upon a more complicated neural model. In other words, we use other models to replace the MF: GMF (He et al. 2017). GMF assigns various weights for different dimensions in the dot-product prediction function, which can be regarded as a generalization of vanilla MF. For GMF, the bridge function directly transforms the user embeddings.

From the results shown in Fig 3, we have several insightful observations: (1) The bridge-based CDR methods can be applied upon various base models. With different base models, both EMCDR, PTUPCDR, and MetaMMCDR effectively improve the recommendation performance for cold-start users in the target domain. Since GMF is a popular and well-designed model in large-scale real-world recommendations, it achieves better performance than vanilla MF. (2) The MetaMMCDR could achieve satisfying performance

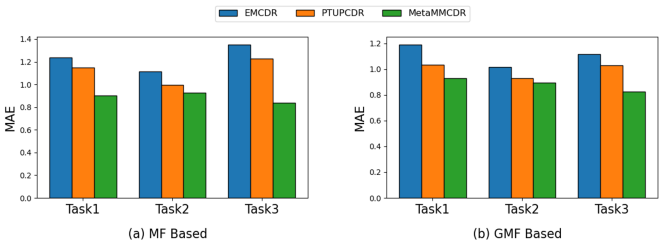

Figure 3: Generalization experiments: applying EMCDR, PTUPCDR and MetaMMCDR upon two base models (a) MF, and (b) GMF.

and the result is good enough to testify the effectiveness of out method in real-world scenarios.

Conclusion

In this study, we investigated cross-domain recommendation (CDR), which aims to transfer user preferences from an auxiliary domain to the target domain. Most existing methods often overlook the rich information contained in multimodal data, such as textual information.

To address this limitation, we proposed a MultiModal Cross-Domain Recommendation framework based on meta-learning methods (MetaMMCDR). Specifically, we leverage text-based embeddings derived from users' historical interaction data to capture personalized preference embeddings. These embeddings are then used as input to train a meta-network that generates personalized bridge functions.

We conducted experiments on real-world datasets to evaluate the effectiveness of the proposed MetaMMCDR framework, and the results demonstrate its efficacy in improving the performance of cross-domain recommendation.

References

Fu, W.; Peng, Z.; Wang, S.; Xu, Y.; and Li, J. 2019. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems. In *AAAI*, volume 33, 94–101.

Hao, X.; Liu, Y.; Xie, R.; Ge, K.; Tang, L.; Zhang, X.; and Lin, L. 2021. Adversarial feature translation for multi-domain recommendation. In *KDD*, 2964–2973.

He, M.; Zhang, J.; Yang, P.; and Yao, K. 2018. Robust transfer learning for cross-domain collaborative filtering using multiple rating patterns approximation. In *WSDM*, 225–233.

He, R.; and McAuley, J. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *AAAI*, volume 30.

He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; and Chua, T.-S. 2017. Neural collaborative filtering. In *WWW*, 173–182.

Hu, G.; Zhang, Y.; and Yang, Q. 2018. Conet: Collaborative cross networks for cross-domain recommendation. In *CIKM*, 667–676.

Kang, S.; Hwang, J.; Lee, D.; and Yu, H. 2019. Semi-supervised learning for cross-domain recommendation to cold-start users. In *CIKM*, 1563–1572.

Lee, H.; Im, J.; Jang, S.; Cho, H.; and Chung, S. 2019. Melu: Meta-learned user preference estimator for cold-start recommendation. In *KDD*, 1073–1082.

Man, T.; Shen, H.; Jin, X.; and Cheng, X. 2017. Cross-domain recommendation: An embedding and mapping approach. In *IJCAI*, volume 17, 2464–2470.

Pan, F.; Li, S.; Ao, X.; Tang, P.; and He, Q. 2019. Warm up cold-start advertisements: Improving ctr predictions via learning to learn id embeddings. In *SIGIR*, 695–704.

Pan, W.; Xiang, E.; Liu, N.; and Yang, Q. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*, volume 24, 230–235.

Singh, A. P.; and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *KDD*, 650–658.

Tao, Z.; Liu, X.; Xia, Y.; Wang, X.; Yang, L.; Huang, X.; and Chua, T.-S. 2022. Self-supervised learning for multimedia recommendation. *IEEE Transactions on Multimedia*, 25: 5107–5116.

Tao, Z.; Wei, Y.; Wang, X.; He, X.; Huang, X.; and Chua, T.-S. 2020. Mgat: Multimodal graph attention network for recommendation. *Information Processing & Management*, 57(5): 102277.

Wang, Q.; Wei, Y.; Yin, J.; Wu, J.; Song, X.; and Nie, L. 2021. Dualgnn: Dual graph neural network for multimedia recommendation. *IEEE Transactions on Multimedia*, 25: 1074–1084.

Xi, D.; Chen, Z.; Yan, P.; Zhang, Y.; Zhu, Y.; Zhuang, F.; and Chen, Y. 2021. Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising. In *KDD*, 3745–3755.

Xie, R.; Liu, Q.; Wang, L.; Liu, S.; Zhang, B.; and Lin, L. 2022. Contrastive cross-domain recommendation in matching. In *KDD*, 4226–4236.

Zhao, C.; Li, C.; Xiao, R.; Deng, H.; and Sun, A. 2020. CATN: Cross-domain recommendation for cold-start users via aspect transfer network. In *SIGIR*, 229–238.

Zhu, F.; Wang, Y.; Chen, C.; Liu, G.; Orgun, M.; and Wu, J. 2018. A deep framework for cross-domain and cross-system recommendations. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, 3711–3717. AAAI. ISBN 9780999241127.

Zhu, Y.; Ge, K.; Zhuang, F.; Xie, R.; Xi, D.; Zhang, X.; Lin, L.; and He, Q. 2021a. Transfer-meta framework for cross-domain recommendation to cold-start users. In *SIGIR*, 1813–1817.

Zhu, Y.; Liu, Y.; Xie, R.; Zhuang, F.; Hao, X.; Ge, K.; Zhang, X.; Lin, L.; and Cao, J. 2021b. Learning to expand audience via meta hybrid experts and critics for recommendation and advertising. In *KDD*, 4005–4013.

Zhu, Y.; Tang, Z.; Liu, Y.; Zhuang, F.; Xie, R.; Zhang, X.; Lin, L.; and He, Q. 2022. Personalized transfer of user preferences for cross-domain recommendation. In *WSDM*, 1507–1515.

Zhu, Y.; Xie, R.; Zhuang, F.; Ge, K.; Sun, Y.; Zhang, X.; Lin, L.; and Cao, J. 2021c. Learning to warm up cold item embeddings for cold-start recommendation with meta scaling and shifting networks. In *SIGIR*, 1167–1176.