IAEAC 2021(ISSN 2689-6621)

Deep Matrix Factorization for Cross-Domain Recommendation

Hailan Kuang[1,2] ,Weiyi Xia[1,2] ,Xiaolin Ma[1,2] ,Xinhua Liu[1,2]
1. School of Information Engineering, Wuhan University of Technology, Wuhan, China
2. Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Wuhan University of Technology, Ministry of Education, Wuhan, China
kuanghailan@whut.edu.cn; weiyixia@whut.edu.cn; maxiaolin0615@whut.edu.cn; liuxinhua@whut.edu.cn
Corresponding Author: Hailan Kuang Email: kuanghailan@whut.edu.cn

2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) | 978-1-7281-8028-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/IAEAC50856.2021.9390866

Abstract—**Recommender system can make personalized recommendation services for all users. Collaborative filtering is the main method can select suitable items to specific users among the recommendation algorithm. Common CF-based approaches use the users' history behaviors which include explicit information or implicit feedback to make recommendation decisions. Matrix Factorization is the most popular idea in the field of collaborative filtering. Currently, many MF-based methods have been proposed and achieve great improvement. But matrix factorization can only fit linear features which limit the their performance in real-world dataset which contains complex and nonlinear feature, moreover, the sparsity of user-item interaction information is another bottleneck of matrix factorization methods. Recently, Deep learning is widely applied to many fields, there are many research has applied deep learning to recommender system. In this paper, we use multi-layer perceptron structures to learn the representation of users and items in ML-based method. On the other hand, in order to address the data sparsity problem exists in collaborative filtering, Cross-Domain Recommendation is a promising way. Combining with collaborative approach to extract the latent feature, we propose Deep Matrix Factorization for Cross Domain Recommendation (DMF-CDR). We test the proposed method on real-world dataset and show that it outperforms several recent popular models.**

Keywords—Recommender system; Matrix Factorization; Cross-domain recommendation; Deep learning

I. INTRODUCTION

Nowadays, with the boosting of online services, people are facing the situation of information overload, for example, some people may have difficulties of choice when encountering large number of items such as news, movies, books and product. Therefore recommendation systems (RSs) play an increasing important role in filtering information for customers. RSs can help people satisfy their interest and explore more items they may need [1]. Now RSs are ubiquitous not only in many websites, such as recommendation of books at Amazon and Douban, movies or videos at Netflix and YouTube, product at Taobao and Jingdong, but also in brick and mortar stores.

Existing approaches for RSs can roughly be categorized into two classes:

(1) Content-based methods [2], they utilize user or item auxiliary information, such as user profile and item description for recommendation.

(2) Collaborative filtering(CF) [3] methods, the main principle of those approaches is based on the past activities or preferences, such as ratings information of users to items.

Generally, because of privacy issue it is difficult to collect user profiles or other user activity. Therefore, *Collaborative filtering* (CF) recommender approaches are widely used in RSs. CF especially matrix factorization (MF) methods are common approaches to recommendation, MF-based methods assume some relationship can be found between users and items by extracting some latent features hidden in the rating matrix. Specifically, MF-based model learns two low-dimensional dense matrix from original sparse rating matrix which represent user and item interests feature, and then use dot product to calculate predicted value.

Recently, MF has become the most popular method to obtain latent factor for recommendation. There are many MF-based method, such as neighbor-based models [4] integrate it with item content, factorization machines [5] can model specific features combination. Most of MF-based methods use dot product to express interaction function, this may model simple and linear features between users and items. This can be a limitation for MF-based models, since there are many complex and nonlinear feature in real-world dataset. Inspired by deep matrix factorization [17], we try to get the non-linear connection between users and items through a deep representation learning structure.

Since Deep Neural Networks(DNNs) has been developed, deep learning methods have shown promising results in various fields such as computer vision [6] and natural language processing [7]. However, the exploration of DNNs on recommender system has received more and more attention.

In fact, every user interact with few items compared to the large number of all items, therefore, rating matrix is very sparse. MF exists data sparsity and cold start problem. Many researchers start use cross-domain recommendation [8] method in order to alleviate above problem. In this work, we use Deep Dual Transfer Cross Domain Recommendation DDTCDR [9] combined with DeepMF [17] to solve above problem which called DMF-CDR.

II. RELATED WORK

A. Recommender Systems

As mentioned above, RSs generate predictions generally based on two classes: CF methods and content-based filtering methods. Content-based filtering methods leverage auxiliary information such as item description, temporal or location information. However this method will face difficulties in collecting those information.

Now CF-based approaches are more popular in recommendation. Matrix Factorization is the main idea for predicting rating value for all users to all items. MF adopts latent vectors to represent users and items in a common low-dimensional space and use dot product to compute the preference score. Many researcher use CF for their task with different data. The most convenient data is *explicit feedback*, which is the rating or preference degree. For example, Douban collects user review history for recommendation. In other way, RSs can refer user preference using *implicit feedback*, which can include purchase or browsing history, search queries or click movements.

MF models extract features from users and items ratings history and get k-dimension vectors. Let $p_i \in \mathbb{R}^k$ denotes user i feature vector, and $q_j \in \mathbb{R}^k$ represents item j feature vector. For user i, the element of p_i measure the extent of interest the user has in items on the corresponding factors, the same as q_j. Let $r_{ui}, \hat{r}_{ui}$ denotes user u 's real value and predicted rating on item i , so the estimate value [11]:

$$\hat{r}_{ui} = \sum_{n=1}^{k} q_{jn}^{T} p_{\text{in}} \tag{1}$$

Conventional MF methods can roughly category into three classes: Eigen Decomposition, Singular Value Decomposition (SVD) and Gradient Descent. The first two algorithms have corresponding faults. Eigen Decomposition can only work on square matrix. SVD algorithm require the rating matrix is dense while rating matrix is very sparse in real-world dataset. Therefore, it is more common to use Gradient Descent to get user vector and item interest vector. The simple principle is shown:

$$\min_{p,q} \sum_{(u,i\in K)}^{n} (r_{ui} - \hat{r}_{ui})^2 \tag{2}$$

B. Deep learning on recommendation

Deep learning is a powerful technique in RSs. For example, deep learning is applied in YouTube recommendation, this system divides the task into two models [12]: Candidate Generation Model and Ranking Model, and explores use MLP for recommendation. Autorec [13] first use auto-encoder (AE) for predicting missing rating value. Then researcher develop more advanced methods like Denoising AE [14], Collaborative Denoising Auto-Encoder (CDAE) [15]. Those AE-based models reconstruct the rating vector from the raw sparse rating vector of user or item. The dense vector which output by encoder in AE structure is effective representation that show the user or item preference and attributes.

NeuMF [16] is framework that use a neural network to replace dot product in conventional MF methods, so the model can be more expressive. The high capacity and nonlinearity of DNN is key for great performance improvement. Moreover, NeuMF presents the general NCF framework, which can combine different model (Generalized Matrix Factorization and Multi-Layer Perceptron are showed in this paper). Deep matrix factorization (DMF) [17] utilize a two pathway structure to map users and items into a low-dimensional space to get a more representative features vectors, which is a simple but feasible way to achieve. Furthermore, DeepCF [11] unifies the two framework that are representation learning and matching function learning under CFNet framework. In a summary, DNN methods are widely used in recommender system, such as substitute for dot product, representation learning and so on. Owing to the non-linearity and expressiveness of DNN, the recommendation system can achieve higher performance.

C. Cross-domain Recommendation

Cross domain recommendation approach is a promising way to deal with the data sparsity problem. This method is extended from single-domain recommendation models. The main principle of this method is exploiting knowledge form auxiliary domains(e.g., movies) which containing auxiliary user preference information to improve recommendation on the target domain(e.g., books) [18].

There are many ways for cross domain recommendation, for example, codebook exemplified by [19], assumes that groups users and items into cluster and matches cluster-level rating pattern across domain. Therefore, the model in this paper can extract a common rating pattern called codebook for knowledge transfer from different domains. In a summary, the central idea of cross domain recommendation is that if one person have the similar preference in different domains, for example, a man like some kinds of movies, we assume that this man have preference for the same kinds of books with high probability. Paper [20] combined cross domain recommendation with Factorization Machine(FM), this is a novel FM-based system which can transfer knowledge from several auxiliary domain to target domain.

Recently, more researchers pay attention to deep learning-based methods. Paper [21] use Domain Separation Network (DSN) [22] for domain adaptation, and utilize a stacked Denoising Autoencoders (SDAE) to reduce difficulty in obtaining a more effective representation. Paper [9] proposed DDTCDR which apply the combination of dual transfer learning mechanism and latent embedding approach to the cross domain recommendation. An orthogonal matrix is used to implement knowledge transfer from auxiliary domains to target domain. Meanwhile, this method consider the duality of the two recommendation models and improve effectiveness of both tasks simultaneously.

III. METHOD

In this section, we present our Deep Matrix Factorization for Cross Domain Recommendation (DMF-CDR). First we construct a deep matrix factorization [17] framework to get the proper representations of users and items. Then we build a knowledge transfer network for cross domain recommendation, which is inspired by paper [9]. The whole architecture of DMF-CDR is showed in Figure 1.

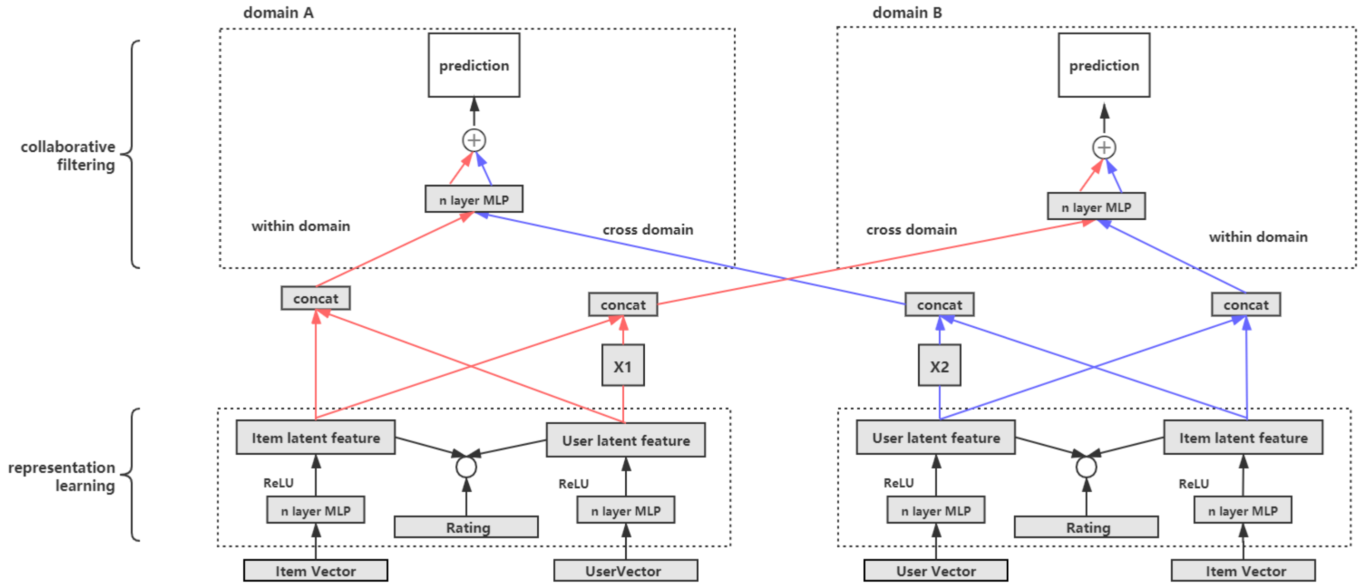

Fig 1: Architecture of DMF-CDR. The architecture is roughly divided into two layer, representation learning network and collaborative filtering network. There are two domains in this architecture, the red and blue line represents the user and item data in domain A and B respectively.

As shown in Figure 1, there are two domain (domain A and domain B). To sum up, the whole process can be divided into two steps: representation learning and collaborative filtering. In the phase of representation learning, two separate MLP structure are used to learning the representations which denote the latent feature of every entity (user or item). In the phase of collaborative filtering, we exploit the representation from the first phase for the collaborative filtering while using two orthogonal matrix for knowledge transfer.

A. Notations

We list the import notations in Table 1.

symbol	Definition
m	The number of users
n	The number of items
$R \in \mathbb{R}^{m \times n}$	The rating matrix
$r_{ij} \in R$	The rating of user i on the item j
R_{i*}	The rating vector of user i
R_{*j}	The rating vector of item j
$\mathbf{p}_i$	The latent factor of user i
$\mathbf{q}_j$	The latent factor of item j
X_1, X_2	The orthogonal metrics for knowledge transfer
$*^A$ and $*^B$	The notations on domains A and B
$\hat{r}_{ij}$	The predicted rating of i on the item j
RS	Recommender system consists of MLP structure

B. Representation Learning

In this step, the network focuses on the representation which denotes the latent feature of all users and items. For m users and n items, we get the whole rating matrix $R \in \mathbb{R}^{m \times n}$. First we get high-dimensional rating vectors R_{i*} (R_{*j}) and feed those vectors into a MLP network to learning corresponding latent factors, then use dot product for predicting ratings. Therefore, the representation learning part for users is defined as:

$$\begin{aligned} \mathbf{a}_0 &= \mathbf{W}_0^T R_{i*} \\ \mathbf{a}_1 &= a(\mathbf{W}_1^T \mathbf{a}_0 + \mathbf{b}_1) \\ &\ldots\ldots \\ \mathbf{p}_i &= \mathbf{a}_N = a(\mathbf{W}_N^T \mathbf{a}_{N-1} + \mathbf{b}_N) \end{aligned} \tag{3}$$

$\mathbf{W}_x$, $\mathbf{b}_x$, $\mathbf{a}_x$ denote the weighed matrix, bias and output in the x-th layer's full-connected layer respectively. $a(\cdot)$ denotes activation function and we use *ReLU* function in this paper. As is shown in (3) we get the latent factor vector $\mathbf{p}_i$ for user i, and get the latent factor vector $\mathbf{q}_j$ for item j in the same way. Generally, the prediction can be calculated in formula (1).

As is shown in (1), we use simple operation dot product to obtain the prediction value $\hat{r}_{ui}$. The reason why we don't utilize more complex structure such as MLP with activation function is that this network focuses on the learning of representation, we can get the non-linear connection between users and items through a deep model, but we can also replace it for a nonlinear structure (MLP) [10].

C. Collaborative Filtering

In this step, as is shown in figure 1, the final predicted ratings in every domain consists of two source, to be more

1

specific, the final prediction in domain A: $\hat{r}^A$ (and domain B: $\hat{r}^B$) is:

$$\begin{aligned}\hat{r}^A &= \alpha RS_A(\mathbf{p}_u^A \oplus \mathbf{q}_j^A) + (1-\alpha) RS_A(X_2^T * \mathbf{p}_u^B \oplus \mathbf{q}_j^B) \\ \hat{r}^B &= \alpha RS_B(\mathbf{p}_u^B \oplus \mathbf{q}_j^B) + (1-\alpha) RS_B(X_1^T * \mathbf{p}_u^A \oplus \mathbf{q}_j^A)\end{aligned} \quad (4)$$

Where α is a hyper-parameter which can be adjusted according to closeness between different domains. We set α=1/2 as default. Another symbol $\oplus$ is concatenation operation.

It is noteworthy that the usage of X_1, X_2. We use two latent orthogonal mapping function for transferring user interests to different domain. Two orthogonal matrix X_1, X_2 to fulfill this task in domain A and domain B because orthogonal matrix preserves similarities between user representation across different latent factor since orthogonal transformation preserves inner product of vectors. Different from paper [9] which use one orthogonal matrix, two orthogonal matrix can accelerate transfer speed. Therefore, those two orthogonal matrix are the key to transfer knowledge between two domains.

IV. EXPERIMENT

In this section, we conduct experiments to validate the effectiveness of our proposed framework.

A. *Experimental Settings*

Dataset We use Amazon Dataset [24] (Amusic and Amovie)1, it is publicly accessible on the websites. On the other hand, Amazon Dataset (Amusic and Amovie) is used for the second step.

Evaluation metric We use *Hit Ratio* (*HR*), *Normalized Discounted Cumulative Gain* (*NDCG*), Root-mean-square error (RMSE) and mean absolute error (*MAE*) as the measure metric. The Calculating Formula is shown as the following:

$$\begin{aligned}HR &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \delta(p_i \le topN) \\ NDCG &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \frac{\log 2}{\log(p_i+1)} \\ RMSE &= \sqrt{\frac{1}{N}\sum_{i,j}(r_{ij}-\hat{r}_{ij})^2} \\ \text{MAE} &= \frac{1}{N}\sum_{i=1}^{n} |r_i - \hat{r}_i|\end{aligned} \quad (5)$$

Where p_i is the hit position for the test item of user i, and $\delta(\cdot)$ is the indicator function. *NDCG* and *HR* evaluate the relevance of recommendation list to user and are both between [0, 1], the more recommendation list relevant to the real user preference, those value are bigger. We use *HR@10* and *NDCG@10* as default setting.

Baseline we compare with following baselines:

- **NCF** [16]. Neural Collaborative Filtering is a deep model that use a multi-layer perception to model latent features of users and items.
- **CoNet** [23]. Collaborative Cross Networks utilizes cross connections to transfer knowledge.
- **DDTCDR** [9]. Deep Dual Transfer Cross Domain Recommendation is a novel approach based on dual learning that transfer knowledge between two domains.

B. *Results*

The result of the first step is shown in Figure 2.

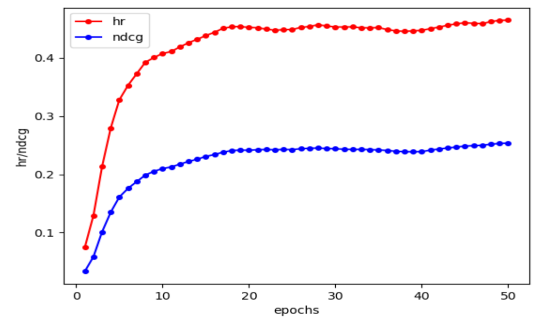

Fig 2. the result of HR and NDCG in representation learning process.

In the first step, we train the model in different layers for many times, and the best performance is shown in Figure 2. This figure illustrates that *HR* and *NDCG* curve tend to be steady after training over 20 epochs.

All the experiment in the second phase results are presented in Table 1. And all models are in the same experimental environment.

TABLE I. THE EXPERIEMTAL RESULTS

method	Evaluation metric			
	RMSE	MAE	HR@10	NDCG@10
NCF	0.8043	0.7302	0.4723	0.2354
CoNet	0.8335	0.7844	0.4745	0.2678
DDTCDR	0.7978	0.7002	0.5067	0.3089
DMF-CDR	**0.7867**	**0.6968**	**0.5111**	**0.3145**

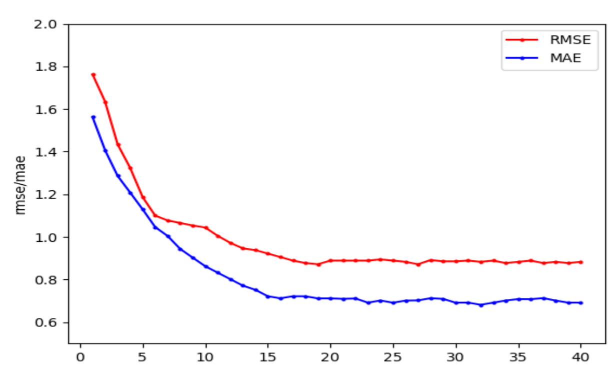

Fig 3. the result of RMSE and MAE of the spare domain

1http://jmcauley.ucsd.edu/data/amazon/

From Table1, we can inference that the proposed method DMF-CDR can provide more accurate predicted ratings because of the minimum value of *RMSE* and *MAE* compared to other methods. On the other hand, *HR@10* and *NDCG@10* shows that DMF-CDR can recommend items which are more satisfactory and relevant items to users. The *RMSE* and *MAE* of the second step is shown in figure 3.

V. Conclusion

In this paper, we proposed a deep model DMF-CDR that has competitive performance. DMF-CDR use MLP networks to learning the representations for users and items which is based on the Deep Matrix Factorization methods and the Dual Transfer Learning used in cross domain recommender. Besides, it is not limited to the formula presented in the paper, this can be extended it to a deep learning structure without the dot product operation.

Experiments demonstrate that the effectiveness of proposed model when facing two domains transfer situation. In the process of training, the performance of two orthogonal matrix has been proved. As a future work, we can find a more effective and efficient way to learning latent factors for users and items for alleviating the data sparsity problem.

Acknowledgment

The work was supported by the National Natural Science F oundation of China (No. 61772088).

References

[1] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet computing, 2003, 7(1): 76-80.

[2] Lang K. Newsweeder: Learning to filter netnews[M]//Machine Learning Proceedings 1995. Morgan Kaufmann, 1995: 331-339.

[3] BadrSarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th international conference on World Wide Web. 2001: 285-295.

[4] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization[C]//Advances in neural information processing systems. 2008: 1257-12

[5] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 2008: 426-434.

[6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[7] Serban I V, Sordoni A, Bengio Y, et al. Building end-to-end dialogue systems using generative hierarchical neural network models[J]. arXiv preprint arXiv:1507.04808, 2015.

[8] Fernández-Tobías I, Cantador I, Kaminskas M, et al. Cross-domain recommender systems: A survey of the state of the art[C]//Spanish conference on information retrieval. sn, 2012: 1-12.

[9] Li P, Tuzhilin A. DDTCDR: Deep Dual Transfer Cross Domain Recommendation[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. 2020: 331-339.

[10] Deng Z H, Huang L, Wang C D, et al. Deepcf: A unified framework of representation learning and matching function learning in recommender system[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 61-68.

[11] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.

[12] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations[C]//Proceedings of the 10th ACM conference on recommender systems. 2016: 191-198.

[13] Sedhain S, Menon A K, Sanner S, et al. Autorec: Autoencoders meet collaborative filtering[C]//Proceedings of the 24th international conference on World Wide Web. 2015: 111-112.

[14] Li S, Kawale J, Fu Y. Deep collaborative filtering via marginalized denoising auto-encoder[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015: 811-820.

[15] Wu Y, DuBois C, Zheng A X, et al. Collaborative denoising auto-encoders for top-n recommender systems[C]//Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. 2016: 153-162.

[16] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]//Proceedings of the 26th international conference on world wide web. 2017: 173-182.

[17] Xue H J, Dai X, Zhang J, et al. Deep Matrix Factorization Models for Recommender Systems[C]//IJCAI. 2017, 17: 3203-3209.

[18] Khan M M, Ibrahim R, Ghani I. Cross domain recommender systems: a systematic literature review[J]. ACM Computing Surveys (CSUR), 2017, 50(3): 1-34.

[19] Li B, Yang Q, Xue X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction[C]//Twenty-First international joint conference on artificial intelligence. 2009.

[20] Loni B, Shi Y, Larson M, et al. Cross-domain collaborative filtering with factorization machines[C]//European conference on information retrieval. Springer, Cham, 2014: 656-661.

[21] Kanagawa H, Kobayashi H, Shimizu N, et al. Cross-domain recommendation via deep domain adaptation[C]//European Conference on Information Retrieval. Springer, Cham, 2019: 20-29.

[22] Bousmalis K, Trigeorgis G, Silberman N, et al. Domain separation networks[J]. arXiv preprint arXiv:1608.06019, 2016.

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