Cross-Domain Sequential Recommendation Based on Self-Attention and Transfer Learning

Dantong Liu1*, Cui Zhu1
1Beijing University of Technology, Chaoyang District, Beijing, 100124, China
*liudt15@emails.bjut.edu.cn

Abstract. Cross-domain recommendation provides effective solutions to the problems of cold start and data sparsity by transferring information from one or more source domains to another target domain. However, people's interactive behaviour usually occurs in a sequence, but most cross-domain recommendation models often ignore the sequential information. At the same time, most sequential recommendation models only consider the case of single-domain. In this paper, we propose a new cross-domain sequential recommendation model (SATLR) based on self-attention and transfer learning, which make recommendations based on user preferences and sequential dependencies learn from cross-domain. The experiment results demonstrate that on the three groups of datasets from amazon, our model is better than other recommendation models.

1. Introduction

With the development of information technology, it is becoming increasingly difficult for people to obtain the information they need from the massive data. In order to solve the problem of information overload, recommender system is proposed [1]. The traditional collaborative filtering mainly based on the user’s historical feedback, but it is difficult to recommend for new users or new items without historical feedback, leading to the problems such as cold start and data sparsity [2]. Cross-domain recommendation has become a solution to these problems [3]. It can transfer information from one or more source domains to another target domain to alleviate the lack of data [4]. Studies have shown that the recommendation effect can be improved by combining information from different domains [5][6].

The items purchased by one user usually come from multiple domains. As shown in figure 1, one user may read several books and also watched a few movies in a certain period of time. User behaviour in different domains is often not independent, and most of them are related to user preferences. In cross-domain recommendation, user preferences are divided into two parts: one part is the user preferences within the domain, and the other part is the cross-domain overlapping user preferences. However, most of the current cross-domain recommendation methods [7][8] focus on the factors such as item features and user preferences, ignoring the information of the user's behaviour sequences.
Figure 1. Schematic diagram of the users’ cross-domain interactions, taking the book domain and the movie domain as examples.

The user’s shopping behaviour usually occurs in a sequence, which means that the next behaviour may depend on the previous behaviour. Sequential recommendation can capture user’s habits, changes in user preferences, etc. Based on user’s historical interactions, sequential recommendation can make the recommendation effect better [9]. The current sequential recommendation methods include [10][11]. But most of the existing models only consider the user's interaction in the same domain, without the connection between different domains. The interactive sequence in other domains may also affect the interactive behaviour of this domain, for example: after watching a movie, the user may continue to read its original novel. Therefore, we consider combining sequential recommendation with cross-domain recommendation to improve the shortcomings that most of the cross-domain recommendation models do not consider the sequential dependencies and most of the sequential recommendation models only consider the single-domain sequences and to obtain better recommendation results based on richer information.

In this paper, we propose a new cross-domain sequential recommendation model (SATLR) based on self-attention and transfer learning, which considers the factor of the interaction sequences in cross-domain recommendation. In addition to user preferences, the sequential dependencies of each domain also participate in the transfer. We first model the sequences of each domain, and then incorporate the information of sequential dependencies into the code of the user’s interaction behaviour. Transfer learning is adopted to model the joint distribution of information in different domains, so as to achieve the purpose of transferring user preferences and sequential dependencies between different domains.

In general, our contributions are summarized as follows:

- We propose a new cross-domain sequential recommendation model. The interaction sequences are integrated into cross-domain recommendation, so that both user preferences and sequential dependencies are involved in transfer learning.
- We integrate cross-domain recommendation and sequential recommendation, retaining the advantages of both. We try to build bridges between the two.
- The experiment results show that our model makes improvement in HR, NDCG and MRR compared to several single-domain and cross-domain recommendation models.

2. Related work

In this section, we provide a brief review of related works. We mainly divide these related works into the following three part.

2.1. Cross-domain recommendation

Generally, cross-domain recommendation can be divided into two categories [4]: one is to aggregate information across different domains [12], and the other is to transfer information from the source domain to the target domain by sharing potential features or rating patterns [13][14].
In order to build more complex cross-domain models, several methods based on deep learning are proposed for cross-domain recommendation. PPGN [15] constructed the preference propagation graph network to propagate user preferences. JSCN [8] uses a joint spectral convolution network to fuse information from multiple domains. BITGCF [16] proposes a novel Bi-directional transfer learning method. In addition, CD-DNN [17] combines the user’s comments in cross-domain recommendation. However, most of the previous studies mainly consider user preferences and item features in the transfer between domains, less considering the user’s interaction sequences.

2.2. Sequential Recommendation
In recent years, sequential recommendation is also one of the research hotspots. It can treat the interactions as dynamic sequences, and transform the sequential dependencies into dynamic user preferences to make better recommendations [18].

For low-order sequential dependencies, Garcin et al. [10] adopt the Markov model. Hidasi et al. [11] use factorization machines for modeling, decomposing the interaction matrix into two low-rank matrices as potential factors. In order to learn high-order sequential dependencies, the main models currently use such as the high-order Markov model [19] and the recurrent neural network model [20], etc. Recently, mainstream sequential recommendation models also often use several sequential modeling techniques which are widely used in the field of natural language processing (NLP) [21][22]. But most of the existing models only consider the user’s interaction history within the domain.

2.3. Cross-domain sequential recommendation
Recently, several studies also attempt to consider both of cross-domain recommendation and sequential recommendation. CDNST [23] uses cross-domain novelty seeking trait for sequential recommendation. According to the user’s historical behaviour, the preference tendency from the perspective of psychology is modeled. SCLSTM [24] adds the rating similarity, feature similarity and time similarity across different domains to the sequences to make sequential recommendations, and the LSTM is improved. These models consider the features of the items, and our model focuses on the interactions between the users and the items.

3. Method
In this section, we formulate the problem and introduce the details about the model SATLR we proposed.

3.1. Problem formulation
In this paper, the cross-domain sequential recommendation problem is first simplified into a triplet of the form \((u, i^a, t^a)\). The same set of users shared by items from two different domains \(X\) and \(Y\) are denoted as \(U\). Therefore, \(u\) in the triplet denotes \(u \in U\{u_1, u_2, ..., u_j\}\), where \(j\) represents the number of users. Which domain the items come from is denoted as \(a\), and the set of items from different domains is denoted as \(I^a\). Therefore, the sets of items from domain \(X\) and \(Y\) are respectively denoted as \(i^X \in I^X=\{i^X_1, i^X_2, ..., i^X_n\}\) and \(i^Y \in I^Y=\{i^Y_1, i^Y_2, ..., i^Y_m\}\), where \(n\) and \(m\) are the numbers of items in domain \(X\) and \(Y\) respectively. The user-item interactions of domain \(X\) and \(Y\) could be respectively represented as the interaction matrices \(R^X \in \mathbb{R}^{j \times n}\) and \(R^Y \in \mathbb{R}^{j \times m}\). The timestamp of the interaction in different domains is denoted as \(t^a\). Therefore, the triplet reflects that in the domain \(a\), the user \(u\) interacts with the item \(i^a\) at the time \(t^a\). \(S\) represents the interaction sequence. According to \(t^a\), we can obtain the sequence of the user \(u\) in each domain as \(S^u \in S=(S^u_1, S^u_2, ..., S^u_b)\), where \(b\) represents the length of the interaction sequence.

3.2. Overview
In this paper, we first model the user's interaction sequences in different domains. This part of the modeling is based on self-attention. The information of sequential dependencies is merged with the user-item interaction information, and it is transferred between the two domains by transfer learning. The encoders and decoders are used to encode and decode the embeddings in the cross-domain
recommendation model to reconstruct the embedding representation so that the new representation contains the information of user preferences and sequential dependencies within the domain and from the other domain. Figure 2 shows the overall framework of the model SATLR we proposed.

In the next parts, we will introduce the details of modeling.

3.3. Sequence modeling

For each domain, we model its own embedding layer. Figure 3 shows the framework of sequence modeling based on self-attention. Taking the domain $X$ as an example, the item embedding matrix of domain $X$ is denoted as $A^X \in \mathbb{R}^{d \times N}$, where $d$ is the potential embedding vector dimension. According to the interaction sequence $S^u$, the input interaction sequence $s = (s_1, s_2, ..., s_N)$ is obtained, where $N$ represents the length of the input sequence in the embedding layer. If the length of $S^u$ is greater than $N$, only the most recent $N$ actions are taken. And if the length of $S^u$ is less than $N$, 0 is added to the left until the length is $N$. In other words, for the user's interaction sequences, only the latest $N$ interactions are considered, and it is thought that the earlier interactions may have little reference value for the sequences. The input embedding matrix is denoted as $B^X \in \mathbb{R}^{N \times d}$.
Self-attention uses global information, so it integrates all relevant interactions into the interaction being processed, without perceiving the position of the interactions in the sequence. Therefore, the position embedding is added to the item embedding, and the new embedding is represented by $\mathbf{B}_\mathbf{z}$.

The scaled dot-product attention [21] is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d}} \right) V$$  \hspace{1cm} (1)$$

where $Q, K$ and $V$ represent the queries, keys and values respectively, and $\sqrt{d}$ is to prevent the inner product value is too large, so that more stable gradient could be made.

The self-attention layer [25] is defined as:

$$H(\mathbf{B}_\mathbf{x}) = \text{Attention}(\mathbf{B}_\mathbf{x}W^0, \mathbf{B}_\mathbf{x}W^K, \mathbf{B}_\mathbf{x}W^V)$$  \hspace{1cm} (2)$$

where $W^0, W^K, W^V \in \mathbb{R}^{d \times d}$ are variable linear matrices, which can be adjusted during model training.

The point-wise two-layer feed-forward network (FFN) [25] is defined as:

$$\text{FFN}(H) = \text{ReLU}(HM_1 + c_1)M_2 + c_2$$  \hspace{1cm} (3)$$

where $M_1, M_2 \in \mathbb{R}^{d \times d}$, and $c_1, c_2$ are $d$-dimensional vectors. As shown in the figure 3, one block composes of a self-attention layer and a feed-forward network, and the block can be stacked multiple times.

3.4. Joint distribution modeling

For the joint distribution modeling, $x_k$ and $y_k$ are used to represent the interaction behaviour of user $u_k$ in domains $X$ and $Y$, which can be regarded as the row vectors of the matrices $R^X$ and $R^Y$. As shown in the figure 2, the outputs of $x_k$ and $y_k$ encoded by the encoders $E^X$ and $E^Y$ are denoted as $z^X_k$ and $z^Y_k$ and the corresponding $f^X_k$ and $f^Y_k$ are spliced, and the new vectors are denoted as $r^X_k$ and $r^Y_k$. When splicing, we provide the variable matrices $\theta^X_k$ and $\theta^Y_k$ for the user $u_k$:

$$r^X_k = [z^X_k \theta^X_k f^X_k] \hspace{1cm} r^Y_k = [z^Y_k \theta^Y_k f^Y_k]$$  \hspace{1cm} (4)$$

$r^X_k$ and $r^Y_k$ contain not only interactive behaviour information that can reflect user preferences, but also interactive sequential information that can reflect sequential dependencies. The transfer of $r^X_k$ and $r^Y_k$ between two domains [26] is defined as:

$$r^{X \rightarrow Y}_k = r^{X}_k W \hspace{1cm} r^{Y \rightarrow X}_k = r^{Y}_k W^T$$  \hspace{1cm} (5)$$

where $W \in \mathbb{R}^{d \times d}$ represents the orthogonal mapping matrix.

The decoder $D^X$ is used to decode $r^X_k$ and $r^{X \rightarrow Y}_k$ into $x_k$, which is the reconstructed $x_k$. Similarly, the decoder $D^Y$ is used to decode $r^Y_k$ and $r^{Y \rightarrow X}_k$ into the reconstructed $y_k$. $E^X, E^Y, D^X$ and $D^Y$ are all two-layer MLP, and ReLU is used as the activation function.

The interactions of the same user $u_k$ in the two domains is represented by a pair of samples $(x_k, y_k)$. $p_\theta(x_k, y_k)$ represents the probability density function [26]:

$$\log p_\theta(x_k, y_k) = D_{KL}[q_\varphi(r^X_k, r^Y_k | x_k, y_k)] + p(r^X_k, r^Y_k) + \text{ELBO}(\theta, \varphi; x_k, y_k)$$  \hspace{1cm} (6)$$

where the first term on the right side of the equation represents the KL divergence of the approximate posterior probability distribution $q_\varphi(r^X_k, r^Y_k | x_k, y_k)$ from the true posterior probability distribution $p(r^X_k, r^Y_k | x_k, y_k)$, and the second term represents the evidence lower bound (ELBO) of $\log p_\theta(x_k, y_k)$:

$$\text{ELBO}(\theta, \varphi; x_k, y_k) = E_{q_\varphi(r^X_k, r^Y_k | x_k, y_k)}[\log p_\theta(x_k, y_k | r^X_k, r^Y_k)]$$

$$- D_{KL}[q_\varphi(r^X_k, r^Y_k | x_k, y_k)]]$$

3.5. Objective function and loss calculation

In this paper, the loss function is divided into three parts: sequence modeling loss, joint reconstruction loss and prior regularization loss. The following is a detailed introduction.

In the first part, according to [25], the binary cross entropy loss is adopted as the objective function:
where $v_{j,t}$ represents the correlation of the $j$-th item regarded as the next interactive item, which is calculated in the prediction layer, according to the output of FFN. $e$ represents the negative item which is randomly generated at each time step in each sequence.

As in equation (7), the first term on the right side of the equation represents the joint reconstruction loss and the second term represents the prior regularization loss of $x_k$ and $y_k$. We make the following proposition: given $x_k$ and $y_k$, $r_k^x$ and $r_k^y$ are conditional independent, given $r_k^x$, $x_k$ and $y_k$ are conditional independent. Therefore, refer to [26], joint reconstruction loss is defined as:

$$
\min_{\varphi^x, \varphi^y, \theta^x, \theta^y, \psi^w} \text{Loss}_{JRL} = -\mathbb{E}_{q_{\varphi^x}(r_k^x|x_k)} \left[ \log p_{\theta^x}(x_k|r_k^x) \right] - \mathbb{E}_{q_{\varphi^y}(r_k^y|y_k)} \left[ \log p_{\theta^y}(y_k|r_k^y) \right] + \lambda \left( \|r_k^x - r_k^x W^T W\|_1 + \|r_k^y - r_k^y W W^T\|_1 \right)
$$

where $\varphi^x, \varphi^y$ are the parameters of encoders, and $\theta^x, \theta^y$ are the parameters of decoders, and $\lambda$ is the weight.

According to the method in [26][27], we let $p(r_k^x, r_k^y) = p(r_k^x)p(r_k^y)$. The adversarial learning is used to represent the prior regularization loss:

$$
\min_{\Phi^x, \Phi^y, \varphi^x, \varphi^y} \max_{\psi^w} \text{Loss}_{PR} = -\mathbb{E}_{r_k^x \sim p(r_k^x)} \left[ \log D^x(r_k^x) \right] - \mathbb{E}_{r_k^y \sim q_{\varphi^x}(r_k^x|x_k)} \left[ \log \left( 1 - D^x(r_k^x) \right) \right]
$$

$$
-\mathbb{E}_{r_k^y \sim q_{\varphi^y}(r_k^y|y_k)} \left[ \log D^y(r_k^y) \right] - \mathbb{E}_{r_k^y \sim q_{\psi^w}(r_k^y|y_k)} \left[ \log \left( 1 - D^y(r_k^y) \right) \right]
$$

where $\Phi^x, \Phi^y$ are the parameters of discriminators $D^x$ and $D^y$. The discriminators are two-layer MLP with ReLU as the activation function, using Gaussian distributions.

The total loss is calculated as:

$$
\text{Loss} = \alpha \text{Loss}_{SAS} + \beta \text{Loss}_{JRL} + \gamma \text{Loss}_{PR}
$$

where $\alpha, \beta, \gamma$ are adjustable parameters, which respectively represent the weight of each loss.

4. Experiments

In this section, we introduce the details of the datasets and the experiments, and give some analysis.

4.1. Dataset

We use the three larger datasets from Amazon: Movies and TV (Movie), Books (Book), CDs and Vinyl (Music). They are grouped in pairs to form three cross-domain datasets for experiment and evaluation: Movie & Book, Book & Music and Movie & Music. In this paper, implicit feedback is used, that is, the user’s ratings as interactions. The rating is an integer from 0 to 5, and the interactions with rating from 3 to 5 are set as positive samples. The overlapping users of the two domains in each group of datasets are selected, and the users who only interact with the single domain are filtered out. We filter out the users and items that interact less than 5 times in one domain. Table 1 shows the statistics of datasets. Density reflects the ratio of the actual number of interactions to the number of possible interactions that all users may have on all items. It can be seen that the data is relatively sparse.

| Datasets               | Users | Domains | Items | Interactions | Density |
|------------------------|-------|---------|-------|--------------|---------|
| Movie & Book           | 29476 | Movie   | 24,091| 591,258      | 0.083%  |
| Book & Music           | 16376 | Book    | 41,884| 579,131      | 0.047%  |
| Movie & Music          | 15914 | Music   | 23,988| 291,325      | 0.074%  |

Table 1. Summary statistics of datasets.
4.2. Baselines
To illustrate the effectiveness, our experiment is compared with several single-domain recommendation methods (CDAE, CFVAE, CMF) and cross-domain recommendation methods (CoNet, DARec, DDTCDR, ETL):
- **CDAE** [28]: It proposes a collaborative denoising auto-encoder for recommendation.
- **CFVAE** [29]: It proposes a multinomial likelihood generative model and Bayesian inference is used to estimate the parameters.
- **CMF** [30]: It uses matrix decomposition to solve the problem of one entity corresponding to multiple relations, sharing parameters among relations.
- **CoNet** [7]: It uses a modified cross neural network to transfer knowledge between different domains.
- **DARec** [31]: It adopts a deep domain adaptation model to learn the invariant user representation from different domains, it can extract and transfer patterns from rating matrices.
- **DDTCDR** [32]: It adopts a dual learning mechanism, using automatic coding technology to construct feature embeddings, and learns cross-domain potential orthogonal mapping to extract user preferences in multiple domains.
- **ETL** [26]: It uses equivalent transformation to capture overlapping attributes between different domains and specific attributes of each domain.

4.3. Experiments and Results
We adjust various parameters and conduct multiple experiments. We set the learning rate \(\in \{0.0008, 0.001, 0.002\}\), hidden units \(\in \{100, 120, 150, 200\}\), the length of the sequence \(\in \{50, 100, 150, 200\}\), the number of blocks in sequence modeling \(\in \{1, 2, 3\}\), the weight \(\lambda \in \{0.5, 1.0, 2.0, 5.0, 6.0\}\), and the weight of each loss \(\alpha \in \{0.5, 1.0, 2.0\}\), \(\beta \in \{0.5, 0.6, 1.0, 2.0\}\), \(\gamma \in \{0.5, 1.0, 2.0\}\).

![Figure 4. Comparison of the results on the dataset of Movie & Book.](image-url)
Figure 5. Comparison of the results on the dataset of Book & Music.

Figure 6. Comparison of the results on the dataset of Movie & Music.

Refer to [26], We randomly select one of the interactions of each user to form the verification set, and one interaction to form the test set, and randomly select 99 non-interactive items as negative samples. We use three evaluation indicators: hit ratio (HR), normalized discounted cumulative gain (NDCG), and mean reciprocal rank (MRR), taking the top5 and top10 of the predicted rankings for experiments. The top5 and top10 of the same indicator are represented by @5 and @10 respectively. For example, HR@5 shows the HR indicator calculated from top5 predicted rankings.

For each pair of datasets, we inspect the recommendation results for each domain separately. Figure 4, 5, 6 show the results of our experiments on the three cross-domain datasets respectively. For the models marked with (*), the results refer to [26]. In the figure 4, 5, 6, each polyline represents the result of one indicator of eight models separately. The results of SATLR we proposed are on the most right of the polylines. According to all the HR, NDCG, and MRR measures, SATLR achieves the best performance on all the three cross-domain datasets.

Through the experiments, when the learning rate is 0.001, the length of the sequence is 200, the number of blocks is 2, $\alpha=1.0$, $\gamma=1.0$, the results of the three sets of experiments are the best. For the dataset of Movie & Book, we set hidden units =150, $\lambda=5.0$, $\beta=1.0$. For the dataset of Book & Music, we set hidden units =120, $\lambda=6.0$, $\beta=0.5$. And for the dataset of Movie & Music, we set hidden units =150, $\lambda=6.0$, $\beta=0.6$.

Compared with the best model ETL, the overall improvement effect of Movie & Book and Book & Music are better. The dataset of Book & Music has the most outstanding performance in the improvement.
of several evaluation indicators. Book & Music has the most improvement in HR for top5 in the book domain, which is 5.12%. Book & Music has the most improvement in NDCG for top5 in the book domain, which is 6.64%, and for top10 also increases by 6.00%. Book & Music has the largest increase in MRR for top5 in the book domain, which is 7.67%, and for top10 also increases by 5.37%.

In general, these results show that cross-domain recommendation is improved greatly by combining the information of sequential dependencies and user preferences.

5. Conclusion

In this paper, we introduce a cross-domain sequential recommendation model based on self-attention and transfer learning. We model the sequence of interactions, adding sequential dependencies to cross-domain recommendation, so as to improve the previous shortcoming that most of the cross-domain recommendations only consider user preferences. We combine user preferences and sequential dependencies to obtain the patterns within the domain and make better recommendations based on the patterns within the domain and learn from other domains.

In our current work, the constructed sequences only consider the sequences of the user’s interactions, but does not consider the time interval between interactions. In the future work, the information of time interval can be tried to add in the sequence modeling to achieve better modeling results. Moreover, the current experiments are only completed on two domains. In the future work, the recommendations may be made across three or more domains.

References

[1] Da’u, A., Salim, N. (2019). Recommendation system based on deep learning methods: a systematic review and new directions. Artificial Intelligence Review, 53: 2709-2748.

[2] Zhao, C., Li, C., Xiao, R., Deng, H., Sun, A. (2020). CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 229-238.

[3] Gao, C., Chen, X., Feng, F., Zhao, K., He, X., Li, Y., Jin, D. (2019). Cross-domain Recommendation Without Sharing User-relevant Data. Proceedings of the World Wide Web Conference. San Francisco. 491-502.

[4] Fernández-Tobías, I., Cantador, I., Kaminaskas, M., Ricci, F. (2012). Cross-domain recommender systems: A survey of the State of the Art. http://arantxa.ii.uam.es/~cantador/doc/2012/7ceri12a.pdf.

[5] Li, B. (2011). Cross-Domain Collaborative Filtering: A Brief Survey. 23rd International Conference on Tools with Artificial Intelligence. Boca Raton. 1085-1086.

[6] Sahebi, S., Brusilovsky, P. (2015). It Takes Two to Tango: An Exploration of Domain Pairs for Cross-Domain Collaborative Filtering. Proceedings of the 9th ACM Conference on Recommender Systems. Vienna. 131-138.

[7] Hu, G., Zhang, Y., Yang, Q. (2018). CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino.667-676.

[8] Liu, Z., Zheng, L., Zhang, J., Han, J., Yu, P.S. (2019). JSCN: Joint Spectral Convolutional Network for Cross Domain Recommendation. 2019 IEEE International Conference on Big Data. Los Angeles. 850-859.

[9] Wang, S., Hu, L., Wang, Y., Cao, L., Sheng, Q.Z., Orgun, M. (2019). Sequential Recommender Systems: Challenges, Progress and Prospects. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao. 6332-6338.

[10] Garcin, F., Dimitrakakis, C., Faltings, B. (2013). Personalized news recommendation with context trees. Proceedings of the 7th ACM conference on Recommender systems. Hong Kong.105-112.

[11] Hidasi, B., Tikk, D. (2015). General factorization framework for context-aware
recommendations. Data Mining and Knowledge Discovery, 30: 342-371.
[12] Sahebi, S., Brusilovsky, P. (2013). Cross-Domain Collaborative Recommendation in a Cold-Start Context: The Impact of User Profile Size on the Quality of Recommendation. 21st International Conference on User Modeling, Adaptation and Personalization. Rome. 289-295.
[13] Tiroshi, A., Kuflik, T. (2012). Domain ranking for cross domain collaborative filtering. 20th International Conference on User Modeling, Adaptation and Personalization. Montreal. 328-333.
[14] Shapira, B., Rokach, L., Freilikhman, S. (2012). Facebook single and cross domain data for recommendation systems. User Modeling and User-Adapted Interaction, 23: 211-247.
[15] Zhao, C., Li, C., Fu, C. (2019). Cross-Domain Recommendation via Preference Propagation GraphNet. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing. 2165-2168.
[16] Liu, M., Li, J., Li, G., Pan, P. (2020). Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. Proceedings of the 29th ACM International Conference on Information and Knowledge Management. 885-894.
[17] Hong, W., Zheng, N., Xiong, Z., Hu, Z. (2020). A Parallel Deep Neural Network Using Reviews and Item Metadata for Cross-Domain Recommendation. IEEE Access, 8: 41774-41783.
[18] Chen, X., Xu, H., Zhang, Y., Tang, J., Cao, Y., Qin, Z., Zha, H. (2018). Sequential Recommendation with User Memory Networks. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey.108-116.
[19] He, R., McAuley, J. (2016). Fusing Similarity Models with Markov Chains for Sparse Sequential Recommendation. 16th IEEE International Conference on Data Mining. Barcelona. 191-200.
[20] Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikík, D. (2016). Session-based Recommendations with Recurrent Neural Networks. 4th International Conference on Learning Representations. San Juan.
[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). Attention is All you Need. 31st Annual Conference on Neural Information Processing Systems. Long Beach. 5999-6009.
[22] Tang, J., Wang, K. (2018). Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. Proceedings of the 11th ACM International Conference on Web Search and Data Mining. Marina Del Rey.565-573.
[23] Zhuang, F., Zhou, Y., Ying, H., Zhang, F., Ao, X., Xie, X., He, Q., Xiong, H. (2020). Sequential Recommendation via Cross-Domain Novelty Seeking Trait Mining. Journal of Computer Science and Technology, 35: 305-319.
[24] Yang, G., Hong, X., Peng, Z., Xu, Y. (2020). Long Short-Term Memory with Sequence Completion for Cross-Domain Sequential Recommendation. 4th Asia-Pacific Web and Web-Age Information Management, Joint Conference on Web and Big Data. Tianjin. 378-393.
[25] Kang, W., McAuley, J. (2018). Self-Attentive Sequential Recommendation. 18th IEEE International Conference on Data Mining. Singapore. 197-206.
[26] Chen, X., Zhang, Y., Tsang, I., Pan, Y., Su, J. (2020). Towards Equivalent Transformation of User Preferences in Cross Domain Recommendation. https://arxiv.org/pdf/2009.06884.pdf.
[27] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I. (2015). Adversarial Autoencoders. https://arxiv.org/pdf/1511.05644.pdf.
[28] Wu, Y., DuBois, C., Zheng, A., Ester, M. (2016). Collaborative Denoising Auto-Encoders for Top-N Recommender Systems. Proceedings of the 9th ACM International Conference on Web Search and Data Mining. San Francisco. 153-162.
[29] Liang, D., Krishnan, R., Hoffman, M., Jebra, T. (2018). Variational Autoencoders for Collaborative Filtering. 27th International World Wide Web. Lyon. 689-698.
[30] Singh, A., Gordon, G.J. (2008). Relational learning via collective matrix factorization. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas. 650-658.
[31] Yuan, F., Yao, L., Benatallah, B. (2019). DARec: Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao. 4227-4233.

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