

Strategies to Solve the Problem of Information Cocoon— Research Progress of Cross-domain Recommendation Algorithm Based on Mining the Potential Interests of Users

Qingyang Bao^{1*} and Dian Tu²

¹*School of Foreign Language, Huazhong University of Science and Technology, Wuhan, China*

²*School of computer and artificial intelligence, Wuhan University of Technology, Wuhan, China*

*Corresponding author: qingyangbao2003@gmail.com

Abstract

The wide application of recommendation algorithms is changing the way people obtain information. There is a potential "information cocoon room effect" in the personalized development of recommendation algorithms, which may restrict the further development of the personalized recommendation system and produce a series of social problems. By consulting the literature and comparing the advantages and disadvantages of several commonly used recommendation algorithms, the cross-domain recommendation system (CDR) proposed in this paper can learn relevant domain knowledge by means of transfer learning to enrich the data sources of the target domain. The problem of the information cocoon room can be solved by mining users' potential interests across domains. This paper focuses on the characteristics and problems of CDR systems, such as mapping-based, transfer-based learning, deep learning and deep neural network learning, in order to provide reference for follow-up research.

Keywords

the Cross-domain Recommender System, Information-Cocoon, Mapping-Based Learning, Transfer-Based Learning

1. Introduction

In the age of explosive growth of information, how to quickly obtain effective information from massive data makes the rapid development of recommendation Algorithms. From obtaining information mainly by the search engine to the extensive application of the recommendation system in recent years, it is changing people's way of obtaining information [1,2].

The optimization of recommender systems in industry is focused on how to improve customer retention, which inevitably ignores the potential "Cocoon effect" and its impact on the audience [2]. The essence of "Information cocoon" is that the current commonly used algorithms cannot accurately obtain the sparse data in the target domain, which restricts the popularization and application of the traditional recommendation system.

The Cross-Domain Recommender System (CDRS) [3,4] considers that there are similarities and correlations between user preferences and project features from different platforms, and it can enrich data by learning assistive domain knowledge to solve the sparsity of target domain data. On the basis of extensive literature review, this paper summarizes the characteristics of several recommendation systems, and summarizes how scholars solve the problem of information cocoon by improving cross-domain recommendation systems in recent years.

2. Research contents

2.1. Overview of the research status of the recommender system

In 1992, Goldberg proposed the first e-mail filtering system, Tapestry. It first proposed a collaborative filtering method, which uses the historical behavior information of users to reorder their emails, so as to filter their emails. The introduction of the Group Lens system in 1994 by the Group Lens group made recommender systems a relatively separate field of study [5]. This system has two

important contributions: First, it brings forward the idea of collaborative filtering to accomplish the recommendation task, and second, it establishes a formal model of the recommendation problem.

Recommendation System(RS) has experienced more than 20 years of development, the current mainstream recommendation algorithm is mainly divided into the following 4 categories.

2.1.1. Rank recommendations based on popularity

According to the popularity ranking, such as Toutiao, Baidu hot search, various film sites, music applications based on popular recommendations. The advantage is that the recommendation is relatively easy to implement and can avoid the cold start of new users. The downside is that the recommendation result is relatively simple and lack of personalization.

2.1.2. Collaborative filtering system

Collaborative filtering system is from the idea: Friends will recommend what they like to each other in real life [6]. At present, there are two popular recommendation algorithms for push collaborative filtering, one is the collaborative filtering of target users and projects, and the other is collaborative filtering based on multiple algorithm models.

There are many examples of successful collaborative filtering, such as the Amazon: the Amazon.com, which generates thirty percentage points of revenue for the Amazon each year. The advantages and disadvantages of collaborative filtering are: 1) the advantages are that little domain knowledge is needed, the model is universal, the engineering implementation is simple, and the effect is very good. 2) The disadvantages are cold start problems (new users/products) ; data sparsity problems; assumptions about past behavior that determine the present without considering the differences in specific scenarios; and hot trends that make it difficult to recommend niche preferences [7].

2.1.3. Content/knowledge based recommendation systems

Content- and knowledge-based recommendations refer users to projects that are similar to their daily favorites, but these similarities are based on the fact that the content and knowledge of the project belong to the same attribute. Advantages: Being able to recommend users' unique minority preferences, to a certain extent, to achieve better recommendations in the case of less data, usually with a fairly good explanation. CONS: it's hard to combine features from different items, and it's hard to surprise the user. If profile mining is not accurate, recommendations are often poor [7].

2.1.4. Hybrid Recommender Systems

Hybrid recommendation algorithm, as its name implies, uses multiple recommendation algorithms to coordinate work in order to overcome the problems existing in a single algorithm and further improve the recommendation effect. Generalized Hybrid recommender systems, including Algorithm mixing, data fusion, multi-scene Behavior Fusion, multi-user points of interest fusion, even user's state continuous change, are generalized hybrid [8].

2.2. Research status of cross-domain Recommendation system (CDR)

Although the recommendation system is in full swing in all walks of life, but because the recommendation content homogeneity problem is serious, causes the user potential interest spot not to receive the value. The CDR uses data-rich domain data to help underserved domains that can make better recommendations to address cold-start or information-cocoon effects, so it has attracted the attention of academia and industry [4].

The general process of CDR: The first step is to obtain background data from the multi-domain. The second step is to input data, that users must communicate with multiple systems to produce

recommended information. The third step is to combine the background data with the input data to come up with the recommended algorithm.

According to the type of method, we classify CDR into three categories:

2.2.1. Mapping based cross-domain recommendation approach

Mapping-based cross-domain recommendation methods model the relationship between two domains through mapping functions. In recent years, several cross-domain recommendation methods based on mapping are described below.

1) *CST(Coordinate System Transfer)*. The idea of CST is to solve the problem of sparse data in the target domain by pre-training user embedding and Item embedding. The author uses the matrix-based transfer learning method to build a model to integrate the user and project content as a subsequent recommended database. The framework establishes a matrix with the auxiliary data, finds the user's initial domain from it, and transfers it to the desired domain, which proves that this method can alleviate the problem of data sparsity. The main framework models are as follows (see Figure 1) [9].

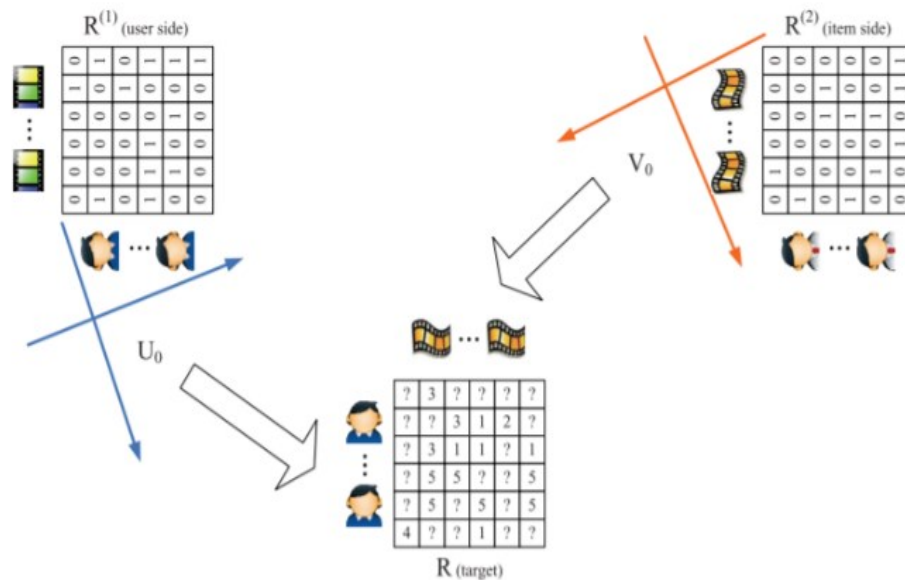


Figure 1: The pattern diagram that CST models between the user and the target [9]

2) *EMCDR (Embedding and Mapping approach for CDR)*. Data sparse is one of the most challenging problems faced by recommendation systems. Tong Man proposes a method to embed and map the EMCDR framework. This framework is different from the traditional cross-domain recommendation model as follows: 1. It uses a multi-layer perception system, which can perceive nonlinear mapping functions. This can provide a more flexible way to learn characteristic content in different fields. 2. By establishing enough databases to achieve the accuracy of the mapping function, to ensure that the sparse data from single-domain sources can also achieve high recommendation efficiency. [10] See Figure 2 for schematic diagram [10].

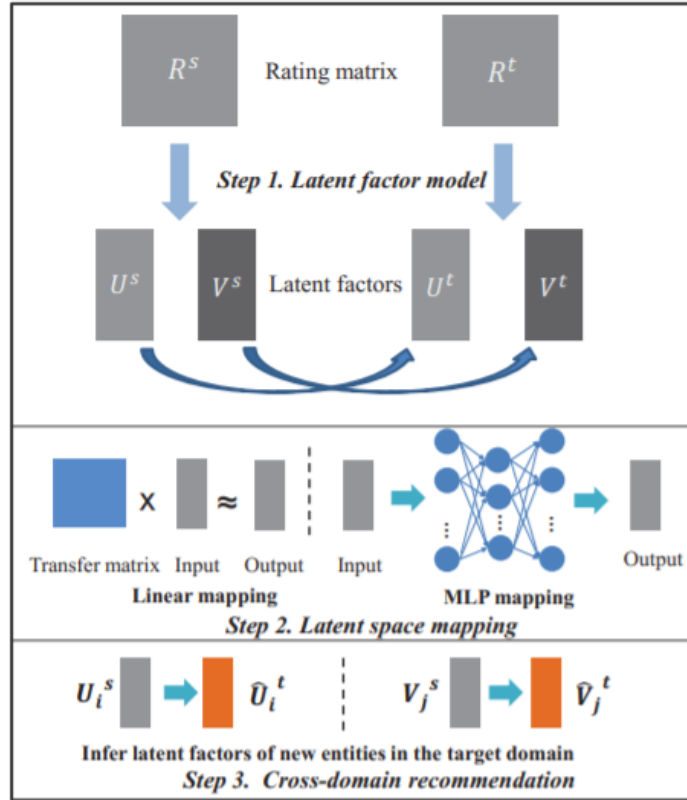


Figure 2: Illustrative diagram of the EMCDC framework [10]

The method first trains a matrix decomposition model in the source domain and obtains the user representation U and the item representation V .

Then the previously shared data is mapped to the target user through the mapping function, and the expression of the mapped data is expected to be close to the user of the target domain.

The implementation of cross-domain recommendation is divided into three steps:

- Build user and item embedding for source and target domains.
- Build the mapping function for the user embedding in the source domain and target domain
- Use the mapping function to represent the user embedding in the source domain as the user embedding in the target domain, and then recall according to the embedding similarity.

3) *SSCDR (Semi-supervised learning for CDR)*. In the recommendation system, EMCDC is an embedded framework designed to calculate the potential vectors of users with sparse data by linear mapping from hidden spaces in another domain. The SSCDR framework can effectively learn cross-domain relationships even with a small amount of tagged data.

The algorithm first learns the vectors of hidden factors of users and projects in each domain, and represents the characteristics of each entity in terms of distance, and then obtains a cross-domain mapping function through training, which marks overlapping users and encodes the distance of each entity by taking non-overlapping items as untagged data[11]. On the basis of this EMCDC, SSCDR also uses an effective reasoning technique to predict the potential vectors of users with less data by aggregating neighborhood information, which is very effective for the expansion of users' potential interests. Finally, in different CDR scenarios, the author has done a large number of experiments. In the real environment where there is a small amount of user overlap between the two domains, SSCDR is better than EMCDC in the accuracy of recommendation [11].

4) *TMCDR (Transfer-Meta Framework for CDR)*. Mapping functions in CDR and EMCDC are only learned on limited overlapping users, and mapping functions tend to be biased towards limited overlapping users, resulting in poor generalization ability. The author establishes a CDR migration meta-framework, including migration phase and meta-phase, which can be applied too many basic models such as MF, PMF, BPF, CML and so on. In order to verify the compatibility of TMCDR, the author

makes extensive experiments on several cross-domain tasks by using the source data of Amazon and Douban. It is proved that TMCDD has strong compatibility, and the recommendation efficiency is better than that of several mapping-based recommendation models mentioned above [12].

2.2.2. Cross-domain recommendation algorithm based on transfer learning

Privacy protection, business competition and other reasons make it difficult for designers of the algorithms to obtain the overlap of user groups in different fields, and accordingly cannot use the overlapping user set as a bridge for the sharing and migration of information resources between domains. The main means to solve these problems is based on collaborative filtering or semantic relation transfer learning.

Transfer learning is widely used in machine learning. It finds the similarity between the original data and the target project through machine learning, then apply the known knowledge to the target domain. This method no longer explicitly models the relationship between the two domains, but trains data from multiple domains directly. This method is very similar to the framework of multi-task learning, and mainly involves knowledge sharing and knowledge transfer in the model structure [13].

1) *CMF(Collective Matrix Factorization)*. Data sparsity is a challenging and universal problem in the recommendation system, which affects the accuracy of recommendation and harms the experience of users. CMF combines the data of the two fields to decompose the matrix and realizes the effect of knowledge transfer through the shared intermediate variable V (which can be shared users or goods).

Zhang et al proposed to acquire knowledge from multiple original databases through machine learning, and then transfer the knowledge to areas where the data is sparse, in order to enrich the sparse data in the target domain, so as to get more effective recommendations. He uses a method to eliminate the deviation between domains, which makes the domains adapt to each other, so that the knowledge of the database will not drift in the process of transfer and learning. Finally, it makes users and projects maintain the consistency of knowledge in the whole process of transfer and learning. In this method, an intermediate subspace is designed, which can extract knowledge from multiple original databases and knowledge from multiple source domains [11].

Some scholars have proposed hierarchical Bayesian model-Collaborative Deep Learning (CDL). This learning model combines content-based deep learning and constructs a matrix through scoring, which not only achieves the effect of collaborative filtering, but also enriches sparse data. In order to verify the recommendation effect of CDL in different fields, the author carries out practical operations on three databases, and the final verification results show that the recommendation efficiency of CDL is very high, especially in the case of less metadata [14].

In order to solve the problem of data sparsity in collaborative filtering. Kuang H [15] and others proposed a deep matrix recommendation algorithm (DMF-CDR) on the basis of traditional CDR. The method is tested on real data sets, and the results show that the performance of this method is better than that of several popular models.

2) *CONET(the collaborative cross networks)*. CDR technology is a valid method to moderate the problem of data rarefaction in the recommendation system according to using knowledge in related fields. Transfer learning is a kind of algorithm behind those technologies. Hu et al. [16] were with the neural network as the basic model, the complex user-project interaction can be learned through deep transfer learning. The author assumes that the veiled layers in the basic networks are linked via mapping to form a synergic network. CONET [16] realizes the cross-domain knowledge transfer by introducing one network to another network and establishing a cross-connection between the two networks. In the multi-layer feedback network, the cooperation between the multi-layer networks is realized by adding interactive links and setting the joint loss function, and the simulation training can be carried out effectively through the back propagation between the multi-layer networks.

Finally, an extensive valuation of the proposed model is running on two large real data sets. Compared with the non-migration method, this model can cut down a large number of drilling examples, but still has the same performance.

3) *PPGN (Preference Propagation Graph Net)*. As a cross-domain method of transfer learning, CONET is indeed better than the non-transfer method, but CONET itself still has some shortcomings, such as inability to capture some higher-order information and difficulty in optimizing its transfer learning model. For these problems, consider using GCN, the PPGN proposed by Zhao et al [17] can solve the limitations of existing methods. The structure of PPGN mainly consists of two components: 1) graph coiling and circulation component; 2) information synthesis and forecast. Zhao et al. used Graph Convolution Network (GCN) to capture some higher-order information. He also constructed a matrix for realizing cross-domain preferences across domains. Cross-domain preference matrix (CDPM) combines user and project interactions in both domains with the advantage of directly disseminating information across different domains. After obtaining the recommendation prediction of the training sample, the prediction performance is improved through joint learning. Finally, the training process is greatly accelerated by splitting, splicing and weighting strategies. Through a large number of experiments, Zhao established a data model based on the preferences of different users in the graph structure, and verified the feasibility and superiority of the PPGN algorithm.

4) *DNN (Deep Neural Networks)*. According to the enhancement of artificial intelligence, recommendation algorithm based on neural network has been used widely in research of traditional recommendation algorithm in recent years. Literature describes the neural network cross-domain recommendation and divides it into two categories [15].

- The cross-domain knowledge recommendation based on neural network

This kind of model mainly utilizes the parameter sharing property between the auxiliary field and the object field in the cross-domain recommendation algorithm to realize the knowledge transfer between the two domains efficiently. [18]

- Cross-domain feature matching based on neural network

This kind of model analyzes the characteristics of objects in the auxiliary field and the object field from a dual perspective and achieves the matching of objects in the two domains directly.

In particular [15], some scholars have proposed a CRD model of the deep neural network, which integrates the above two types of neural network cross-domain recommendation by using the sharing layer of neural network. The specific process is as follows: firstly, we extract characteristic data from auxiliary domain and target domain, then these characteristics are compounded through the sharing layer. Finally, the first N-bit recommendation of the project is realized through the function of the score prediction layer. He [19] proposed that the learning of this model was realized by the neural collaborative filtering model (NCF).

5) *KerKT(kernel-induced Knowledge Transfer)*. At present, most mainstream cross-domain recommendation systems are generally based on the assumption that elements in different fields either overlap completely or do not overlap at all. However, in practice, the partial overlap of recommended entities in different areas is more common, and the overlapping parts may still have different ways of expression in different areas. To solve these problems, some scholars have proposed a cross-domain recommendation system (KerKT) based on kernel-induced knowledge transfer to solve the problem that partial overlap affects the accuracy of cross-domain recommendation. The strength of KerKT lies in its ability to handle partially overlapping entities and to correct biases in cross-domain recommendations in practice [20]. The steps are as follows (see Figure 3):

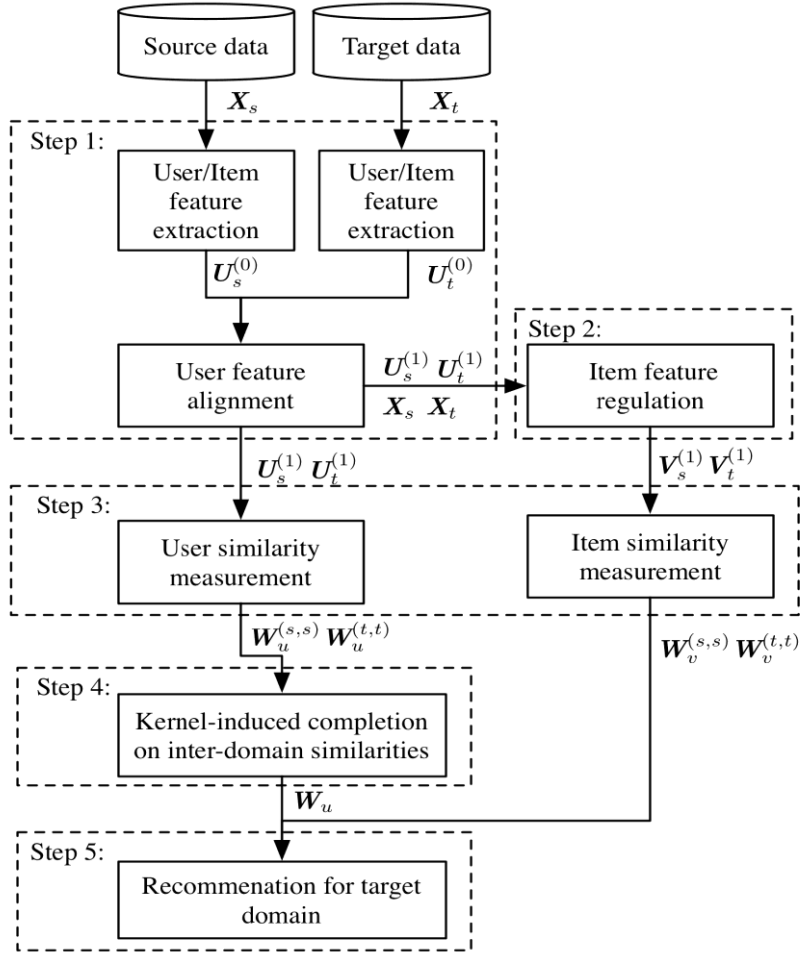


Figure 3: The procedure of the KerKT method [20]

- Step1. Extract user features and project features $U_s^{(0)}$ and $U_t^{(0)}$ of the source domain and target domain, respectively according to the original scoring matrix X_s and X_t , and then align the two groups of user features to the same feature space through the user overlap to obtain $U_s^{(1)}$ and $U_t^{(1)}$.
- Step2. Adjust the project features according to the original scoring matrix X_s X_t and the aligned user feature matrix $U_s^{(1)}$ $U_t^{(1)}$, i.e. $V_s^{(1)}$ $V_t^{(1)}$.
- Step3. Use $U_s^{(1)}$ $U_t^{(1)}$ $V_s^{(1)}$ $V_t^{(1)}$ obtained in the first two steps to measure the similarity between user and item in a domain to obtain $W_u^{(s,s)}$ $W_u^{(t,t)}$ $W_v^{(s,s)}$ $W_v^{(t,t)}$
- Step4. Use kernel-induced completion to measure the user similarity between domains and obtain W_u
- Step5. The user/item features are retrained based on the constraints of The entity whose position is obvious (CMF is used here); then, recommendations are made [21].

3. Conclusion

This paper first introduces the development of the recommendation system, then analyzes the advantages and disadvantages of the traditional single domain recommendation algorithm, and mainly introduces the recommendation algorithm that can effectively alleviate the information cocoon effect to a certain extent -- CDR. In this paper, cross-domain recommendation algorithms are divided into cross-domain recommendation based on mapping and multi-domain collaborative filtering recommendation based on transfer learning. The cross-domain recommendation algorithms proposed by different researchers are summarized and sorted out, and the differences of different cross-domain recommendation algorithms in implementation difficulty and prediction accuracy are compared. With the booming development of artificial intelligence industry, it can be predicted that recommendation

algorithm technology will be more widely integrated with the deep learning model in the future, which undoubtedly opens up a bright road to better solve the problem of information cocoon house.

4. References

- [1] Alexander Felfernig MJ, Gerald Ninaus, Florian Reinfrank and Stefan Reiterer. (2005) Toward the Next Generation of Recommender Systems: Applications and Research Challenges. *IEEE Transactions on Knowledge and Data Engineering* ., 17(6):734-49.
- [2] Wenqian Y. (2021) Research and Analysis on the Construction of Network Ideology under We Media. 2021 2nd International Conference on Artificial Intelligence and Education (ICAIE) ., pp. 345-8.
- [3] Li B. (2011) Cross-Domain Collaborative Filtering: A Brief Survey. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence., pp. 1085-6.
- [4] Cremonesi P, Tripodi A, Turrin R. (2011) Cross-Domain Recommender Systems. 2011 IEEE 11th International Conference on Data Mining Workshops. , pp. 496-503.
- [5] Resnick PIN, Suchak M, Bergstrom P, Riedl J. (1994) GroupLens: An open architecture for collaborative filtering of Netnews. *Proc Cscw* ., 10:175-86.
- [6] Yang WPQ. (2013) Transfer Learning in Collaborative Filtering for Sparsity Reduction. *Artificial Intelligence* ., 197:39–55.
- [7] Greg Linden, Brent Smith, and Jeremy York. (2003) Amazon.com_recommendations_item-to-item_collaborative_filtering. *IEEE Internet Computing*., 7(1):76-80.
- [8] Burke R.(2002)Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*., 12:331-70.
- [9] Weike Pan EWX, Nathan N. Liu and Qiang Yang. (2010) Transfer Learning in Collaborative Filtering for Sparsity Reduction. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*., 230-35.
- [10] Man T, Shen H, Jin X, Cheng X. (2017) Cross-Domain Recommendation: An Embedding and Mapping Approach.,10: 175-186
- [11] Kang S, Hwang J, Lee D, Yu H.(2019)Semi-Supervised Learning for Cross-Domain Recommendation to Cold-Start Users. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China: Association for Computing Machinery., pp. 1563–72.
- [12] Zhu Y, Ge K, Zhuang F, Xie R, Xi D, Zhang X, Lin L, He Q.(2021)Transfer-Meta Framework for Cross-domain Recommendation to Cold-Start Users. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* ., pp. 1813-7.
- [13] Hao P, Zhang G, Martinez L, Lu J. (2019) Regularizing Knowledge Transfer in Recommendation with Tag-Inferred Correlation. *IEEE Trans Cybern.*, 49:83-96.
- [14] Elkahky A, Song Y, He X, Acm. (2015)A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. 24th International Conference on World Wide Web (WWW). Florence, ITALY., pp. 278-88.
- [15] Kuang H, Xia W, Ma X, Liu X.(2021)Deep Matrix Factorization for Cross-Domain Recommendation. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)., pp. 2171-5.
- [16] 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*., pp.667-76.
- [17] Zhao C, Li C, Fu C. (2019) Cross-Domain Recommendation via Preference Propagation GraphNet. *The 28th ACM International Conference* ., 2165-68
- [18] He, X., Liao, L., Zhang, H., Nie, L., & Chua, T. S. (2017).Neural Collaborative Filtering. *The 26th International Conference. International World Wide Web Conferences Steering Committee*. 173–182 <https://doi.org/10.1145/3038912.3052569>
- [19] Chauhan S, Mangrola R, Viji D.(2021)Analysis of Intelligent movie recommender system from facial expression. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)., pp. 1454-61.

- [20] Qian Z, Jie L, Wu D, Zhang G. (2018) A Cross-Domain Recommender System With Kernel-Induced Knowledge Transfer for Overlapping Entities. IEEE Transactions on Neural Networks and Learning Systems ., 1-15.
- [21] Zhang Q, Lu J, Wu D, Zhang G. (2019) A Cross-Domain Recommender System With Kernel-Induced Knowledge Transfer for Overlapping Entities. IEEE Trans Neural Net Learn Syst., 30:1998-2012.