Exploiting Aesthetic Preference in Deep Cross Networks for Cross-domain Recommendation

Jian Liu
Institute of AI, Soochow University
Suzhou, China
ppzhao.suda.edu.cn

Pengpeng Zhao*
Institute of AI, Soochow University
Suzhou, China
ppzhao.suda.edu.cn

Fuzhen Zhuang†
University of Chinese Academy of Science(CAS)
Beijing, China
zhuangfuzhen@ict.ac.cn

Yanchi Liu
Rutgers University
New Jersey, USA
yanchi.liu@rutgers.edu

Victor S. Sheng
Texas Tech University
Texas, USA
victor.sheng@ttu.edu

Jiajie Xu
Soochow University
Suzhou, China
xujj@suda.edu.cn

Xiaofang Zhou
The University of Queensland
Queensland, Australia
zxf@itee.uq.edu.au

Hui Xiong
Rutgers University New Jersey
New Jersey, USA
hxiong@rutgers.edu

ABSTRACT

Visual aesthetics of products plays an important role in the decision process when purchasing appearance-first products, e.g., clothes. Indeed, user's aesthetic preference, which serves as a personality trait and a basic requirement, is domain independent and could be used as a bridge between domains for knowledge transfer. However, existing work has rarely considered the aesthetic information in product images for cross-domain recommendation. To this end, in this paper, we propose a new deep Aesthetic Cross-Domain Networks (ACDN), in which parameters characterizing personal aesthetic preferences are shared across networks to transfer knowledge between domains. Specifically, we first leverage an aesthetic network to extract aesthetic features. Then, we integrate these features into a cross-domain network to transfer users' domain independent aesthetic preferences. Moreover, network cross-connections are introduced to enable dual knowledge transfer across domains. Finally, the experimental results on real-world datasets show that our proposed model ACDN outperforms benchmark methods in terms of recommendation accuracy.

CCS CONCEPTS

• **Information systems → Recommender system**.

KEYWORDS

Cross-domain Recommendation;Knowledge Transfer;Aesthetic Feature

ACM Reference Format:
Jian Liu, Pengpeng Zhao, Fuzhen Zhuang, Yanchi Liu, Victor S. Sheng, Jiajie Xu, Xiaofang Zhou, and Hui Xiong. 2020. Exploiting Aesthetic Preference in Deep Cross Networks for Cross-domain Recommendation. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380036

1 INTRODUCTION

Recommender systems have attracted a great amount of interests in recent years. They are utilized to handle the information overload problem and help people make right decisions according to their historical behaviors. When shopping online, we usually look through product images before making the decision, especially for products that are important in appearance, e.g., clothes, shoes. Product images provide abundant visual information, including design, color schemes, decorative patterns, texture, and so on. We can even estimate the quality and the authenticity of a product from its images. As a result, visual information could play an important role in improving the performance of recommending products with appearance priority.

Researchers have started to use image data for recommendation with various image features, such as features extracted by convolutional neural networks, the scale-invariant feature transform algorithm, and color histograms [3, 24–26]. These image features contain semantic information to distinguish items and have been proved effective in recommendation tasks. One important visual factor, aesthetics, has rarely been considered in previous visual content enhanced recommender systems. If you know more about your consumer's aesthetic preferences, you can recommend the products more convincingly according to consumer's taste. However, few efforts have been found considering the aesthetic preferences for

*The Corresponding author
†Fuzhen Zhuang is also with Key Lab of Intelligent Information Processsing of CAS

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380036

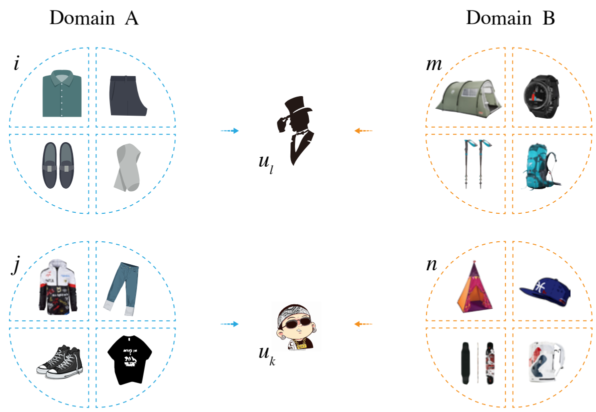


Figure 1: A user's aesthetic behaviors in different areas have consistency.

recommendation except Yu et al. [22] introduce aesthetic information into clothing recommender systems. They demonstrate that incorporating aesthetic features can improve the recommendation performance significantly. However, it does not consider aesthetic features for cross-domain recommendation.

Nowadays users are active in various E-commerce websites and have generated a large number of behavioral data in different domains. Although the aesthetic preference varies significantly from user to user, a user's aesthetic behaviors in different domains could be consistent. For example, as shown in Figure 1, if a user u_l likes the simple black and white style, she/he will prefer item i in domain A and item m in domain B. If a user u_k likes bells and whistles, a hip hop style, she/he will prefer item j in domain A and item n in domain B. Based on the above observation, we can see the aesthetic behavioral data of domain A may help model the aesthetic preferences in domain B. This consistency of aesthetic behavior is helpful for cross-domain recommendation, especially when one domain suffers from the data sparsity issue.

To capture aesthetic preferences and to transfer knowledge among different domains, we propose the new deep Aesthetic Cross-Domain Networks, termed as ACDN, in which parameters characterizing the personal aesthetic preferences are shared across different domains to improve recommendation. Specifically, we first leverage an aesthetic network to extract the holistic features to represent the aesthetic elements of a product photo (for example, the aesthetic elements can be color, structure, proportion, style, etc.). Then, we incorporate the aesthetic features into a deep cross-domain network to transfer users' domain independent aesthetic preferences. Moreover, dual knowledge transfer, which can enable the domains benefit from each other, is achieved by using cross-connections and joint loss function. Finally, we conduct extensive experiments to evaluate the effectiveness of the proposed model ACDN on two real-world Amazon datasets. The experimental results show that ACDN achieves better performance in terms of the ranking metric, comparing with various baselines. We further conduct a thorough analysis to understand how the aesthetic features and transferred knowledge help improve the performance of ACDN. In summary, the main contributions of this paper are listed as follows:

- To the best of our knowledge, we are the first to leverage novel aesthetic features for cross-domain recommendation by capturing users' domain independent aesthetic preferences. Moreover, we compare the effectiveness of the aesthetic features with different types of conventional features for cross-domain recommendation to demonstrate the advantage of the aesthetic features.
- We propose a new cross-domain recommendation model ACDN for better modeling an individual's propensity from the aesthetic perspective for recommendation, where the aesthetic preference of each individual is shared for knowledge transfer across different domains.
- We conduct extensive experiments on two real-world cross-domain datasets. Our experimental results show the proposed model ACDN outperforms the state-of-the-art methods via comprehensive analysis.

2 RELATED WORK

Recommender system is usually seen as predicting users' preferences on unobserved items based on their past history interactions. Collaborative filtering (CF) is an early popular and widely used recommendation method based on matching users with similar tastes or interests [6]. One representative technology for CF is Matrix Factorization (MF), which learns latent factors of users and items from a user-item rating matrix [10, 15]. However, these CF-based methods based on the sole rating matrix are faced with data sparse and the cold-start problem.

Cross-Domain Recommendation (CDR) [2] is an effective technique for alleviating data sparse issues by leveraging the rating information from other domains to enhance the performance on the target domain [23]. Existing CDR methods can be divided into two groups, i.e., content-based and transfer-based. Berkovsky et al. [1] proposed a content-based CDR approach targeting the data sparsity problem by importing and aggregating vectors of users' ratings operating in different application domains. Later on, Winoto et al. [21] uncovers the association between user preferences on related items across domains. Transfer-based approaches mainly employ machine learning techniques (e.g., transfer learning and neural networks) to transfer knowledge across domains. Li et al. [11] proposed a codebook method, which transfers user-item rating patterns from an auxiliary task in other domains to a sparse rating matrix in a target domain. Man et al. [13] proposed an embedding and mapping framework (EMCDR), which uses a multi-layer perceptron to learn the nonlinear mapping function between a source domain and a target domain. In terms of neural network, Misra et al. [14] proposed a convolutional network with cross-stitch units to learn an optimal combination of shared and task-specific representation using multi-task learning, and hence enable the knowledge transfer between two domains. However, these methods treat knowledge transfer as a global process with shared global parameters and do not match source items with the specific target item given a user. Different from the above works, we introduce novel aesthetic features for cross-domain recommendation to capture users' domain independent aesthetic preference and propose a new deep aesthetic preference cross-domain network for better modeling an individual's propensity from the aesthetic perspective for recommendation.

3 NOTATIONS AND PROBLEM DEFINITION

In this section, we will introduce related notations and our problem settings. Given a target domain $\mathcal{T}$ and a source domain $\mathcal{S}$, where users $\mathcal{U}$ (its size $m = |\mathcal{U}|$) are shared, we want to transfer knowledge across domains. We denote the set of items in source domain $\mathcal{S}$ as I_S and the size of items in source domain is $n_S = |I_S|$. Similarly, we denote the set of items in target domain $\mathcal{T}$ as I_T and its size is $n_T = |I_T|$. We use u to index a user, i to index a target item and j to index a source item. Then, matrix $\boldsymbol{R}_T \in \mathbb{R}^{m \times n_T}$ is used to represent the user-item interaction matrix in the target domain, and the entry $r_{ui} \in \{0, 1\}$ is 1 if the user u has purchased the item i and 0 otherwise. Similarly for the source domain, matrix $\boldsymbol{R}_S \in \mathbb{R}^{m \times n_S}$ is used to describe user-item interactions, the entry $r_{uj} \in \{0, 1\}$ is 1 if user u has an interaction with item j and 0 otherwise. Here each domain can be treated as a problem of collaborative filtering for implicit feedback [8, 17].

For the task of item recommendation, our goal is to recommend a ranked list of items for each user based on his/her history records, i.e., top-N recommendation. We aim to improve the recommendation performance in the target domain with the help of the user-item interaction information and user's aesthetic preference from the source domain. The items are ranked by their predicted scores: $\hat{r}_{ui} = f(u, i|\Theta)$, where f is an interaction function and Θ are model parameters. For matrix factorization techniques, the match function is the fixed dot product: $\hat{r}_{ui} = \boldsymbol{P}_u^T \boldsymbol{Q}_i$, and parameters $\Theta = \{\boldsymbol{P}, \boldsymbol{Q}\}$ are latent vectors of users and items, where $\boldsymbol{P} \in \mathbb{R}^{m \times d}$, $\boldsymbol{Q} \in \mathbb{R}^{n \times d}$ and d is the dimension size. For neural CF approaches, neural networks are used to a parameterized function f and learn it from interactions: $f(\boldsymbol{x}_{ui}|\boldsymbol{P}, \boldsymbol{Q}, \theta_f) = \phi_o(\phi_L(...(\phi_1(\boldsymbol{x}_{ui}))))$, where the input $\boldsymbol{x}_{ui} = [\boldsymbol{P}^T \mathbb{X}_u, \boldsymbol{Q}^T \mathbb{X}_i]$ is merged from projections of the user and the item, and the projections are based on their one-hot encodings $\mathbb{X}_u \in \{0, 1\}^m, \mathbb{X}_i \in \{0, 1\}^n$ and embedding matrices $\boldsymbol{P} \in \mathbb{R}^{m \times d}$, $\boldsymbol{Q} \in \mathbb{R}^{n \times d}$. The output and the hidden layers are computed by ϕ_o and ϕ_l ($l \in [1, L]$) in a multi-layer feedforward neural network (FFNN), and the connection weight matrices and biases are denoted by θ_f. In our aesthetic preference cross-domain recommendation network, each domain is modeled by a neural network, and these networks are jointly learned to improve the performance through mutual knowledge transfer.

4 THE PROPOSED MODEL

In this section, we briefly describe the proposed Aesthetic preference Cross-Domain Network model (ACDN). As is shown in Figure 2(a), ACSR consists of three components: a Embedding Layer, a Aesthetic Feature Extraction, and a cross Transfer Layer. In the following subsections, we will introduce each part of our model and corresponding network learning strategy in detail.

4.1 Aesthetic Feature Extraction

We utilize the pre-trained deep aesthetic neural network ILGNet [9] to extract aesthetic features from item images. ILGNet (I : Inception, L : Local, G : Global) is a novel deep convolutional neural network , which introduces the inception module into image aesthetics classification and can extract aesthetic features from low level to high level. As is shown in Figure 2(b), this network connects the layer of local features to the layer of global features to form a concat layer of 1024 dimension, which are binary patterns. Specifically, the first and the second inception layers are considered to extract local image features and the last inception layer is considered to extract global image features after two max pooling and one average pooling. Then, we connect the output of the first two inception layers (256 dimension for each) and the last inception layer (512 dimension) to form a 1024 dimension concat layer as the holistic aesthetic feature.

In our work, for each item i in the target domain, we utilize the pre-trained ILGNet to extract its aesthetic features $\boldsymbol{x}_i^a \in \mathbb{R}^{1 \times 1024}$ from the corresponding image in advance. Similarly, for each item j in the source domain, we obtain its aesthetic feature $\boldsymbol{x}_j^a \in \mathbb{R}^{1 \times 1024}$. With the aesthetic features of items, we can capture users' aesthetic preference across domains and improve the target domain recommendation performance.

4.2 Embedding Layer

To represent the input, we encode user-item interaction indices by one-hot encoding. For user u, item i from the target domain and item j from the source domain, we map them into one-hot encoding: $\mathbb{X}_u \in \{0, 1\}^m, \mathbb{X}_i \in \{0, 1\}^{n_T}, \mathbb{X}_j \in \{0, 1\}^{n_S}$ where only the element corresponding to index is 1 and others are 0. Then, we embed one-hot encodings into continuous representation: $\mathbf{x}_u = \boldsymbol{P}^T \mathbb{X}_u, \mathbf{x}_i = \boldsymbol{Q}_t^T \mathbb{X}_i, \mathbf{x}_j = \boldsymbol{Q}_s^T \mathbb{X}_j$ To get high level representation, we use a multi-layer perceptron (MLP) to get the hidden representation: $x_u = g(\boldsymbol{W_u}\ \mathbf{x}_u + \boldsymbol{b_u}), x_i = g(\boldsymbol{W_i}\ \mathbf{x}_i + \boldsymbol{b_i}), x_j = g(\boldsymbol{W_j}\ \mathbf{x}_j + \boldsymbol{b_j})$.

Then we concatenate $\boldsymbol{x}_i' = [\boldsymbol{x}_i, \boldsymbol{x}_i^a]$ and $\boldsymbol{x}_j' = [\boldsymbol{x}_j, \boldsymbol{x}_j^a]$, where $\boldsymbol{x}_i^a$ is the aesthetic feature of item i and $\boldsymbol{x}_j^a$ is the aesthetic feature of item j. Finally, we concatenate $\boldsymbol{x}_{ui} = [\boldsymbol{x}_u, \boldsymbol{x}_i, \boldsymbol{x}_i^a]$, $\boldsymbol{x}_{uj} = [\boldsymbol{x}_u, \boldsymbol{x}_j, \boldsymbol{x}_j^a]$ to be the input of following building blocks.

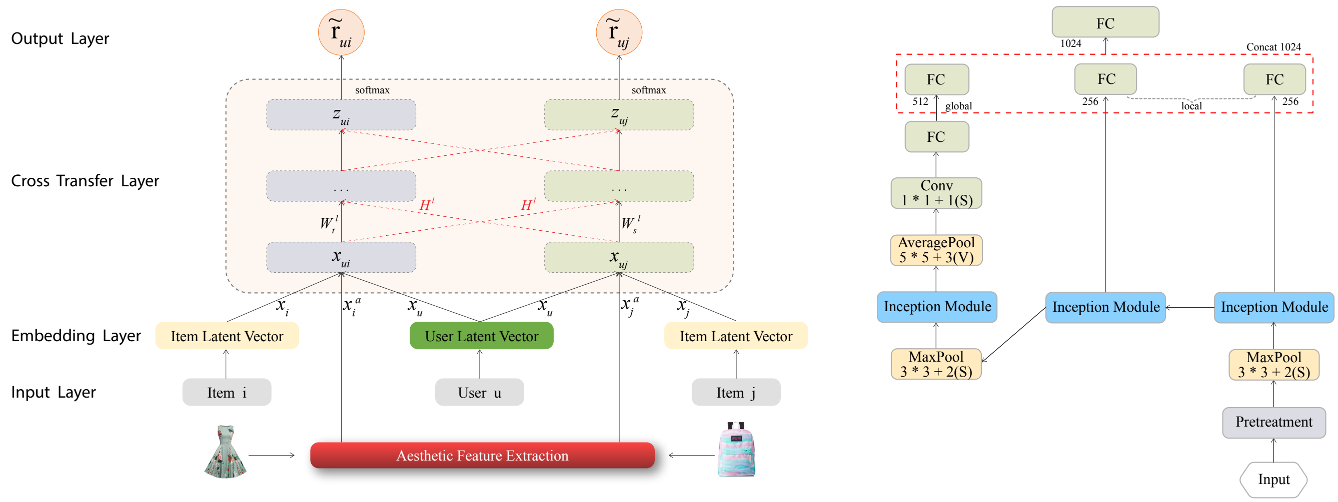
4.3 Cross Transfer Layer

In this subsection, we will introduce the cross transfer layer for knowledge transfer in detail. Different from CSN [14], the core idea of the cross transfer unit is to adopt a relationship/transfer matrix rather than a scalar weight to transfer knowledge. The target domain can receive information from the source domain and vice versa. As is shown in Figure 2(a), we add cross transfer units to the entire FFNN. Denote $\boldsymbol{W}_t^l$ as the weight connecting from the l-th layer to the $(l+1)$-th layer and b_t^l as the bias in target domain. Similarly, there are $\boldsymbol{W}_s^l$ and b_s^l in the source domain. Denote $\boldsymbol{H}^l$ as the relationship matrix from the l-th layer to the $(l+1)$-th layer. The two base networks can be coupled by cross transfer unit:

$$\alpha_t^{l+1} = \sigma(\boldsymbol{W}_t^l \alpha_t^l + \boldsymbol{b}_t^l + \boldsymbol{H}^l \alpha_s^l), \tag{1}$$

$$\alpha_s^{l+1} = \sigma(\boldsymbol{W}_s^l \alpha_s^l + \boldsymbol{b}_s^l + \boldsymbol{H}^l \alpha_t^l), \tag{2}$$

where σ is the activation function and we use ReLU [16] here. In the target domain, we can observe that the representations of the $(l+1)$-th layer α_t^{l+1} receives two information flows: one is from the transform gate controlled by a weight matrix $\boldsymbol{W}_t^l$ and another is from transfer gate controlled by $\boldsymbol{H}^l$ (similarly for the α_s^{l+1} in the source domain). This way of knowledge transfer happens in two



(a) The proposed deep aesthetic preference cross-domain network architecture.

(b) The aesthetic network (ILGNet) architecture.

Figure 2: The left figure is the proposed deep aesthetic preference cross-domain model architecture and the right figure is the aesthetic network (ILGNet) architecture.

directions, from the source domain to the target domain and from the target domain to the source domain, which can enable dual knowledge transfer across domains and let them benefit from each other. Similar to CSN [14], we take the same relationship/transfer matrix $\boldsymbol{H}^l$ for both directions to reduce model parameters and make the model compact. Actually, it does not improve the performance of recommendation by taking different transfer matrices for two directions.

Obviously, the relationship/transfer matrix $\boldsymbol{H}^l$ is very crucial to our model. We assume that not all representations from another domain are useful and we expect that the representations receiving from other domains are selective and useful. This corresponds to enforcing a sparse prior on the structure and can be achieved by penalizing the relationship/transfer matrix $\boldsymbol{H}^l$ via regularization. We take the widely used sparsity-induced regularization: least absolute shrinkage and selection operator [20]. We enforce the l_1-norm regularization on the relationship/transfer matrix $\boldsymbol{H}^l$ to induce sparsity:

$$\Omega(\boldsymbol{H}^l) = \lambda \sum_{i=1}^{r} \sum_{j=1}^{q} |h_{i,j}|, \tag{3}$$

where h_{ij} is the entry (i, j) of $\boldsymbol{H}^l$, hyper-parameter λ controls the degree of sparsity and $r \times q$ is the size of matrix $\boldsymbol{H}^l$. It means that $\boldsymbol{H}^l$ linearly transforms representations $\alpha_s^l \in \mathbb{R}^q$ in the source domain and the result is as part of the input to the next layer $\alpha_t^{l+1} \in \mathbb{R}^r$ in the target domain.

Since we focus on one-class collaborative filtering, the output is the probability that the input pair is a positive interaction. This can be achieved by a softmax layer:

$$\hat{r}_{ui} = \phi_o(z_{ui}) = \frac{1}{1 + exp(-\boldsymbol{h}^T z_{ui})}, \tag{4}$$

where $\boldsymbol{h}$ is a model parameter.

4.4 Model Learning

According to the task of item recommendation and the nature of the implicit feedback, we adopt cross-entropy as our loss function for model optimization. The objective function to be minimized in the model optimization is defined as follows:

$$\mathcal{L}_0 = - \sum_{(u,i) \in \mathbf{R}^+ \cup \mathbf{R}^-} r_{ui} log \hat{r}_{ui} + (1 - r_{ui}) log(1 - \hat{r}_{ui}), \tag{5}$$

where $\mathbf{R}^+$ and $\mathbf{R}^-$ are the observed interaction matrix and randomly sampled negative examples [17], respectively. This objective function has probabilistic interpretation and is the negative logarithm likelihood of the following likelihood function:

$$\mathcal{L}(\Theta | \mathbf{R}^+ \cup \mathbf{R}^-) = \prod_{(u,i) \in \mathbf{R}^+} \hat{r}_{ui} \prod_{(u,i) \in \mathbf{R}^-} (1 - \hat{r}_{ui}), \tag{6}$$

where Θ are model parameters.

we add *joint loss function* to our proposed model, which can be trained efficiently by back-propagation. Instantiating the base loss $\mathcal{L}_0$ described in Eq.5 by the loss of the target domain ($\mathcal{L}_T$) and loss of the source domain ($\mathcal{L}_s$), the objective function of our proposed model is their joint losses:

$$\mathcal{L}(\Theta) = \mathcal{L}_t(\Theta_t) + \mathcal{L}_s(\Theta_s), \tag{7}$$

where the model parameters $\Theta = \Theta_t \cup \Theta_s$. This objective function can be optimized by stochastic gradient descent (SGD):

$$\Theta' \leftarrow \Theta - \eta \frac{\partial \mathcal{L}(\Theta)}{\partial \Theta}, \tag{8}$$

where η is the learning rate.

5 EXPERIMENTS

In this section, we will introduce the experiment settings and make comprehensive analysis on experimental results.

Table 1: Performance Comparison of different methods on two datasets. The best performance is highlighted in boldface.

Dataset	Clothing & Home Improvement (Dataset 1)									Outdoor and Sports & Clothing (Dataset 2)								
Method	*TopN = 5*			*TopN = 10*			*TopN = 20*			*TopN = 5*			*TopN = 10*			*TopN = 20*		
	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR
BPRMF	0.0902	0.0753	0.0650	0.1730	0.0941	0.0704	0.2757	0.0939	0.0785	0.1105	0.0853	0.0766	0.1743	0.0975	0.0779	0.2848	0.1371	0.0892
VBPR	0.1027	0.0831	0.0778	0.1836	0.1103	0.0811	0.2903	0.1142	0.0831	0.1335	0.0970	0.0885	0.1976	0.1105	0.0861	0.3044	0.1501	0.1023
CMF	0.1201	0.0903	0.0812	0.2014	0.1213	0.0947	0.3189	0.1242	0.0811	0.1479	0.1022	0.0931	0.2214	0.1233	0.1005	0.3237	0.1601	0.1255
CDCF	0.1130	0.0863	0.0794	0.1904	0.1178	0.0803	0.3054	0.1183	0.0877	0.1338	0.0928	0.0876	0.2103	0.1167	0.0931	0.3155	0.1566	0.1148
MLP	0.1251	0.0926	0.0866	0.2079	0.1225	0.0871	0.3266	0.1385	0.0988	0.1533	0.1047	0.0958	0.2321	0.1280	0.1021	0.3304	0.1622	0.1295
MLP++	0.1292	0.0957	0.0974	0.2101	0.1278	0.0944	0.3321	0.1379	0.1033	0.1590	0.1136	0.1011	0.2467	0.1339	0.1104	0.3367	0.1734	0.1356
CSN	0.1388	0.1022	0.0922	0.2179	0.1335	0.1027	0.3465	0.1424	0.1104	0.1655	0.1243	0.1033	0.2498	0.1449	0.1170	0.3390	0.1881	0.1408
CoNet	0.1437	0.1059	0.1014	0.2230	0.1383	0.1143	0.3524	0.1513	0.1185	0.1739	0.1328	0.1124	0.2539	0.1480	0.1241	0.3437	0.1938	0.1510
ACDN	**0.1472**	**0.1077**	**0.1047**	**0.2289**	**0.1403**	**0.1166**	**0.3601**	**0.1560**	**0.1220**	**0.1763**	**0.1357**	**0.1140**	**0.2611**	**0.1529**	**0.1254**	**0.3529**	**0.2003**	**0.1543**
Improve	2.4%	1.69%	3.2%	2.6%	1.44%	2.95%	2.18%	3.10%	2.01%	1.38%	2.18%	1.40%	2.83%	3.31%	1.04%	2.68%	3.35%	2.18%

Table 2: Dataset Description

Dataset	Statistics	Source Domain Clothing	Target Domain Home Improvement
Dataset 1	#user	8673	8673
	#item	21317	18442
	#interactions	60942	56183
	#density	0.032%	0.035%
Dataset	Statistics	Source Domain Outdoor and Sports	Target Domain Clothing
Dataset 2	#user	13164	13164
	#item	17765	22465
	#interactions	68291	82416
	#density	0.029%	0.029%

5.1 Experimental Setup

5.1.1 Dataset. We study the effectiveness of our proposed approach on a real-world public dataset *Amazon*[*] with different kinds of domains. It contains product reviews and metadata from Amazon, which has been used to evaluate the performance of various approaches. Here we use three domains: *Home Improvement, Clothing, Outdoor and Sports,* and conduct experiments on two datasets with following combinations. The basic statistics of datasets are listed in Table 2.

5.1.2 Evaluation Protocol. For the item recommendation task, the leave-one-out evaluation is widely used and we follow the protocol in [5]. It means that we reserve one interaction as the test item for each user. We determine hyper-parameters by randomly sampling another interaction per user as the validation set. We follow the common strategy which randomly samples 99 negative items that are not interacted by the user and then evaluate how well the recommender can rank the test item against these negative ones. Since we aim at $TopN$ item recommendation, the typical evaluation metrics are hit ratio (HR), normalized discounted cumulative gain (NDCG) and mean reciprocal rank (MRR), where the ranked list is cut off at topN = $\{5, 10, 20\}$.

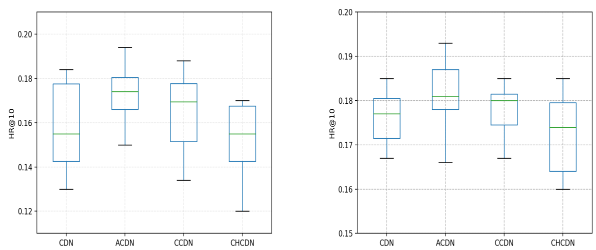
- **BPRMF**: Bayesian personalized ranking [18] is a typical collaborate filtering approach, which learns the user and item latent factors via matrix factorization and pairwise rank loss.
- **MLP**: Multi-layer perception [5] is a neural collaborate filtering approach, which can learn a user-item interaction function by neural networks.
- **MLP++**: We combine two MLPs by sharing the user embedding matrix. This is a degenerated method that no cross transfer units.
- **VBPR**: VBPR [4] is a scalable factorization model to incorporate visual signals into predictors of people's opinions, which can make use of visual features extracted from product images by pre-trained deep networks.
- **CDCF**: Cross domain Collaborate Filtering [12] with factorization machines is a state-of-the-art cross-domain recommendation which extends FM. It is a context-aware approach which applies factorization on the merged domains (aligned by the shared users).
- **CMF**: Collective matrix factorization [19] is a multi-relation learning approach, which jointly factorizes matrices of individual domains. Here, the relation is user-item interaction. The shared user factors enable knowledge transfer between two domains.
- **CSN**: The cross-stitch network [14] is a deep multitask learning model and jointly learns two base networks. It enables knowledge transfer by a linear combination of activation maps from two domains via a shared coefficient.
- **CoNet**: CoNet [7] is the latest collaborative cross networks for cross-domain recommendation, which can enable dual knowledge transfer across domains by introducing cross connections from one base network to another and vice versa and let them benefit from each other.

[*] http://jmcauley.ucsd.edu/data/amazon/links.html

5.2 Performance Comparison

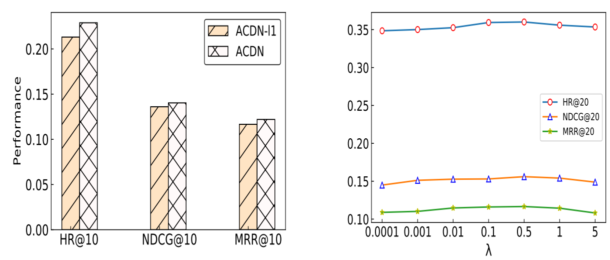
The results of different methods are illustrated in Table 1. It can be observed that:

Firstly, we can find that cross-domain methods (i.e., CMF and CDCF) produce a better performance than single-domain methods (i.e., BPRMF and VBPR) at all settings on both datasets, regardless of shadow methods and deep methods. This indicates that cross-domain methods benefit from knowledge transfer and is an effective



(a) Dataset 1 (b) Dataset 2

Figure 3: (a)Necessity of the Aesthetic Features. (b)The impact of the l_1-norm regularization for Sparsity.



(a) l_1-norm Regularization (b) Sensitivity of λ

Figure 4: (a):Impact of l_1-norm Regularization.(b):Sensitivity Analysis of λ.

technique for alleviating the data sparsity issue. VBPR outperforms BPRMF, which indicates that visual features extract from item images can indeed enhance the performance of recommendation.

Secondly, we can notice that deep methods perform better than shadow methods in both single-domain and cross-domain. For example, MLP improves more than 15% comparing with shadow methods BRPMF and VBPR in all cases in single-domain, and deep cross-domain models (i.e., MLP++, CoNet, and CSN) outperform shadow cross-domain models (i.e., CMF and CDCF) in all cases on two datasets. This shows the effectiveness of deep neural models with the non-linear combination and more parameters can benefit not only single-domain recommendation but also cross-domain recommendation.

Thirdly, we can observe that our proposed neural model ACDN is better than all baselines on both two datasets at each setting, including the base MLP network, shallow cross-domain models (i.e., CMF and CDCF), deep cross-domain models (i.e., MLP++, CoNet, and CSN). These results demonstrate the effectiveness of the proposed aesthetic features enhanced the cross-domain neural model.

Finally, comparing MLP++ and MLP, sharing user embedding is slightly better than the base network due to unilateral knowledge transfer, which shows the necessity of dual knowledge transfer in a deep way. CSN is inferior to CoNet on both datasets. The reason is possible that the assumption of CSN is not appropriate: all representations from the auxiliary domain are equally important and are all useful. This motivates us to learn what to transfer adaptively and filter irrelevant information for target domain recommendation by using a cross transfer matrix rather than a scalar weight. Also, our model outperforms the state-of-the-art method CoNet since CoNet merely transfers user-item rating information, which demonstrates that aesthetic features can help improve cross-domain recommendation performance, especially in appearance-first products.

In summary, the empirical comparison results demonstrate the superiority of the proposed neural model to transfer aesthetic preference and source domain knowledge for cross-domain recommendation.

5.3 Advantage of the Aesthetic Features

We combine various widely used features in our basic model and compare the effect of each type of features by constructing models: **CDN**(Remove the aesthetic features), **CHCDN**(Replace the aesthetic features with color histograms) and **CCDN**(Replace the aesthetic features with CNN features). Figure 3 shows the distribution of 10 maximum at HR@10 on both Datasets during 40 iterations. We can observe that CHCDN performs the worst since the low-level features are too crude and unilateral, and can provide very limited information about consumers' aesthetic preference for cross-domain. Our model ACDN, with aesthetic information, performs the best on both datasets, though CNN features also contain some aesthetic information (like color, texture, etc.). It is far from a comprehensive description, which can be provided by the aesthetic features on account of the abundant raw aesthetic features inputted and training for knowledge transfer for cross-domain recommendation. CNN features can perform better than aesthetic features in a single domain [22], but experiments demonstrate the effectiveness of the aesthetic features in cross-domain recommendation. This phenomenon proves our assumption that a user's aesthetic preference is domain independent and can be used as a bridge between domains for knowledge transfer.

5.4 Impact of Hype-Parameters

Figure 4(a) shows the impact of l_1-norm regularization on the entries h_{ij} of $\boldsymbol{H}^l$ in Eq.3 . ACDN-l_1 is that we remove the l_1-norm regularization from our model. From the experimental results, we can observe that ACDN performs better than ACDN-l_1 on Dataset 1, which demonstrates the effectiveness of enforcing the sparse structure (l_1-norm regularization) on the cross transfer matrices. The l_1-norm regularization can control knowledge transfer between source domain and target domain. In other words, with l_1-norm regularization, our model can utilize the cross transfer matrices to select representations adaptively to transfer for cross-domain recommendation.

We also analyze the sensitivity of the penalty parameter λ of l_1-norm regularization and we optimize the performance of our model varying with $\lambda \in (0.001, 0.01, 0.1, 0.5, 1, 5)$. As is shown in Figure 4(b), our model achieves the best performance with setting λ=0.5 on Dataset 1.

6 CONCLUSIONS

In this paper, a new deep Aesthetic preference Cross-Domain Network (ACDN) was introduced to transfer users' aesthetic preferences across different domains to enhance the recommendation performance. Specifically, we proposed a deep cross-domain recommendation network incorporated with aesthetic preferences, which enabled dual knowledge transfer across domains by introducing cross transfer unit from one base network to another. The experimental results show that our model significantly outperforms the state-of-the-art approaches on cross domain recommendation.

ACKNOWLEDGMENTS

This research was partially supported by NSFC (No.61876117, 61876217, 61872258, 61728205), Open Program of Key Lab of IIP of CAS (No. IIP2019-1) and PAPD.

REFERENCES

[1] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2007. Cross-domain mediation in collaborative filtering. In *ICUM*. Springer, 355–359.

[2] Paolo Cremonesi, Antonio Tripodi, and Roberto Turrin. 2011. Cross-domain recommender systems. In *2011 IEEE 11th International Conference on Data Mining Workshops*. IEEE, 496–503.

[3] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI, 144–150.

[4] Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *Thirtieth AAAI Conference on Artificial Intelligence*. AAAI, 144–150.

[5] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.

[6] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. 1999. An algorithmic framework for performing collaborative filtering. In *SIGIR*. ACM, 230–237.

[7] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 667–676.

[8] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 263–272.

[9] Xin Jin, Le Wu, Xiaodong Li, Xiaokun Zhang, Jingying Chi, Siwei Peng, Shiming Ge, Geng Zhao, and Shuying Li. 2019. ILGNet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation. *IET Computer Vision* 13, 2 (2019), 206–212.

[10] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* (2009), 30–37.

[11] Bin Li, Yang Qiang, and Xiangyang Xue. 2009. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. In *IJCAI*. IJCAI.

[12] Babak Loni, Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Cross-domain collaborative filtering with factorization machines. In *European conference on information retrieval*. Springer, 656–661.

[13] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *IJCAI*. IJCAI, 2464–2470.

[14] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. 2016. Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 3994–4003. https://doi.org/10.1109/CVPR.2016.433

[15] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *NIPS*. Curran Associates, Inc., 1257–1264.

[16] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.

[17] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. 2008. One-class collaborative filtering. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 502–511.

[18] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI, 452–461.

[19] Ajit P Singh and Geoffrey J Gordon. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 650–658.

[20] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* (1996), 267–288.

[21] Pinata Winoto and Tiffany Tang. 2008. If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations. *New Generation Computing* 26, 3 (2008), 209–225.

[22] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. 2018. Aesthetic-based clothing recommendation. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 649–658.

[23] Yu Zhang, Bin Cao, and Dit Yan Yeung. 2012. Multi-Domain Collaborative Filtering. *UAI* (2012), 725–732.

[24] Lili Zhao, Zhongqi Lu, Sinno Jialin Plan, and Qiang Yang. 2016. Matrix Factorization+ for Movie Recommendation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI, 3945–3951.

[25] Pengpeng Zhao, Chengfeng Xu, Yanchi Liu, Victor S Sheng, Kai Zheng, Hui Xiong, and Xiaofang Zhou. 2019. Photo2trip: Exploiting visual contents in geo-tagged photos for personalized tour recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[26] Pengpeng Zhao, Xiefeng Xu, Yanchi Liu, Victor S. Sheng, Kai Zheng, and Hui Xiong. 2017. Photo2Trip: Exploiting Visual Contents in Geo-tagged Photos for Personalized Tour Recommendation. In *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*. ACM, 916–924.