Transfer Learning via Linking and Generation for Distant Sparse Cross-domain Recommendation

ABSTRACT

In this paper, we study a novel cross-domain recommendation problem termed distant sparse cross-domain recommendation (DSCR). In DSCR, the majority of users in the source and target domains are non-overlapped, and the user interaction behavior is very sparse. Most traditional cross-domain recommendation methods can only be applied to overlapped users in two domains, and their experimental datasets are denser than DSCR. The problems to be solved by these methods are completely different from DSCR, which makes them unable to handle DSCR well.

As a solution, we propose linking and generation framework (LGF). The proposed framework expresses the correlation between the two domains into a cross-domain bipartite graph based on the long-term and short-term interests of users. The graph is the information basis for knowledge transfer between domains. It avoids the problem of being easily affected by heterogeneous information due to extremely sparseness. Furthermore, we are the first to propose domain mutual information (DMI) to evaluate the amount of information contained in the cross-domain bipartite graph, and its value indicates the correlation between the two domains and can guide the construction of the cross-domain bipartite graph. Then, LGF adopts multi-task generative network (MTG) generates behavior in the source domain for target domain non-overlapped users. This approach can expand the scope of application of the cross-domain recommendation methods. Finally, LGF uses the adaptive knowledge transfer network (AKTN) to mine source domain information that is more complementary to the target domain to maximize the effect of knowledge transfer. Experimental offline results and online A/B tests in recommendations of DiDiChuXing show that LGF outperforms state-of-the-art algorithms.

KEYWORDS

cross-domain recommendation, sparse, linking, generation

ACM Reference Format:

. 2018. Transfer Learning via Linking and Generation for Distant Sparse Cross-domain Recommendation. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

As one of the most popular online serving platforms of immediate ride-hailing in the world, DiDiChuxing serves hundreds of millions of active passengers and millions of active drivers. For different requirements of daily operation, DiDiChuxing builds a dozen different types of recommendation scenarios, such as coupon recommendation, product recommendation, etc. However, there is a common problem in these recommendation scenarios. The data of these scenarios is very sparse, which makes it difficult to train a satisfactory model with the data of a single scenario. Utilizing the data from other scenarios to improve the recommendation performance of this scenario is an intuitive solution.

In recent years, cross-domain recommendation[3] has attracted the attention of many researchers. Cross-domain recommendation uses the information in the source domain to improve the effectiveness of the recommendation task in the target domain. For example, transfer information from Twitter to YouTube[17, 19], and from Electronics to Cell Phones[12]. However, despite the growing success of cross-domain recommendation methods, traditional methods can only be applied to overlapped users who exist in both domains, and the remaining non-overlapped users cannot benefit from these methods. In the recommendation scenarios of DiDiChuxing, there are relatively few overlapped users between domains. Even if a large effect is achieved on the data of overlapped users, there is not much improvement in the overall scope. In addition, compared with scenarios of DiDiChuxing, the user interaction data in the experiments of the traditional methods are denser. For example, in CATN[25] and BiTGCF[12], users with interaction numbers below 10 and 5 in the data set are filtered out. However, in the scenarios of DiDiChuxing, the average number of user interactions within the confidence period is less than 2. In the case of extremely sparse data, mixing any heterogeneous information in the process of knowledge transfer may lead to negative transfer[16]. For example, when a user has 100 interactive items, passing him an incorrect item information will not have much impact on him. However, if a user has only 2 interactive items, any wrong information passed to him will cause him great damage. Traditional methods, no matter sharing user embedding[1, 20] or user preference mapping[4, 13, 22], are essentially linking two domains based on users. We abstract this method of linking domains as Figure 1a. In the case of sparse user behavior, the type of linking domain method cannot well represent the correlation between domains, and may even cause negative transfer due to any noise introduced in the process of knowledge transfer. The sparsity of user behavior in existing methods is completely different from that in DSCR, which makes it difficult to obtain good results even on overlapped users. The low proportion of overlapped users and the extreme sparse data make it difficult for existing methods to work in the scenarios of DiDiChuxing.

In this paper, we focus on the cross-domain recommendation problem with a small proportion of overlapped users and extremely sparse data. We denote this problem as Distant Sparse Cross-domain Recommendation(DSCR). In this problem, the proportion of overlapped users in the target domain users is less than 15%, and the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

average number of interactions of users in source domain and target domain is less than 2.

To solve DSCR, we propose the LGF framework. The LGF framework builds a cross-domain bipartite graph based on the long-term(e.g.,age group, gender, city and other profile features) and short-term(e.g., items the user clicked) interests of users. The linking domain method of LGF is shown in Figure 1b. The cross-domain bipartite graph describes the correlation between domains and is the basis of knowledge transfer. The amount of information contained in the graph is positively correlated with the correlation between domains. Compared with the existing linking domain method, the cross-domain bipartite graph better represents the correlation between domains. Then, we propose domain mutual information(DMI) to measure the correlation between two domains. DMI evaluates the information content in the cross-domain bipartite graph from two aspects: difference and diversity. This is the first measurement method to measure correlation between domains in cross-domain recommendation. DMI can further guide the construction of cross-domain bipartite graphs, thereby enhancing the effect of knowledge transfer.

In DSCR, a larger portion of users in the target domain are non-overlapped users, and generating information on the source domain for them is a solution to improve the recommendation effect. Generative cross-domain methods e.g. CnGAN[18], use a generative model to generate information in the source domain for target domain non-overlapped users. Their training data is the data of overlapped users, as shown in Figure 2a, the test data is the data of non-overlapped users, and the data distribution of the training set and the test set are exactly the same. However, the proportion of overlapped users in DSCR is so small that it is difficult to train a satisfactory generator. And in real industrial scenarios, the data distribution of overlapped users and non-overlapped users is inconsistent, and the generator may perform poorly on non-overlapped users, and may even cause negative transfer due to the generation of heterogeneous information. LGF proposes the multi-task generative network (MTG) to generate source domain behavior for the target domain non-overlapped users. MTG consists of a source domain generation model and a target domain classification model. The source domain generation model makes full use of the source domain data for training, as shown in Figure 2b, avoiding the shortcomings of insufficient training of generative cross-domain methods due to insufficient training data. The target domain model of MTG is the regularization of the source domain generative model to prevent it from overfitting to the source domain data.

The essence of cross-domain recommendation is to solve the problem of insufficient information in a single domain based on data fusion. When the information provided by the input sources represents different parts of the scene and could thus be used to obtain more complete global information [2]. LGF proposes an adaptive knowledge transfer network (AKTN) to perform cross-domain knowledge transfer. The core idea of AKTN is to increase the weight of the source domain items that differ greatly from the target domain information, and reduce the weight of source domain items that are similar to the target domain information. AKTN effectively reduces the influence of redundant information in the source domain on the results.

The main contributions of this work are summarized in the following:

1) We are the first to propose distant sparse cross-domain recommendation(DSCR). In DSCR, the proportion of overlapped users in the target domain users is less than 15%, and the average number of interactions of users in source domain and target domain is less than 2.

2) We build a cross-domain bipartite graph based on the long-term interests and short-term interests of users, which describes the correlation between the two domains and is the basis for cross-domain knowledge transfer. It improves the tolerance of cross-domain models to heterogeneous information in the case of sparse data. In order to measure the amount of information in the cross-domain bipartite graph, we propose domain mutual information (DHI). To the best of our knowledge, graph mutual information is the first measurement method to measure correlation between domains in the cross-domain recommendation.

3) We propose multi-task generative network (MTG) to generate source domain behavior for target domain non-overlapped users. MTG makes full use of the source domain data, and effectively reduces the loss caused by the inconsistent distribution of the training set and the test set.

4) In order to maximize the use of information in the source domain, we propose an adaptive knowledge transfer network (AKTN) to mine information that is more auxiliary to the target domain.

5) We conduct extensive offline experiments and deploy online A/B tests at DiDiChuxing. Experimental results show the superiority of our LGF over state-of-the-art algorithms.

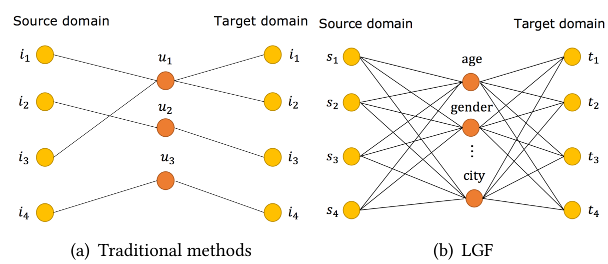
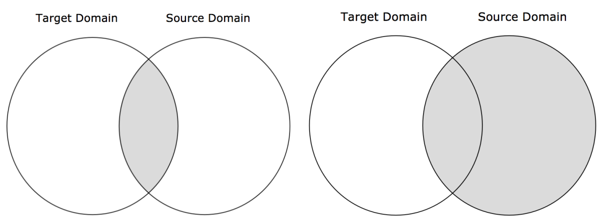


Figure 1: Comparison of linking domains between Traditional methods and LGF

2 RELATED WORK

Cross-domain recommendation uses the relevant source domain as auxiliary information to solve the data sparseness and cold start problems of the target domain. CMF[20] method extends Matrix Factorization by jointly learning user embeddings and item embeddings from multiple rating matrices. CDTF[7] uses tensor factorization to capture the triadic relation of user-item-domain. SocialMF[9] solves the problem of social recommendation by introducing friend relations, and its core approach is to restrict the user's embedding to be similar to his friends. CROSS[11] solves cross-platform recommendation by combining SocialMF and CMF.



(a) Generative cross-domain methods (b) MTG

Figure 2: Comparison of trainning data between Generative cross-domain methods and MTG

Table 1: Notations and descriptions used in this paper

Notation	Descriptition
S, T	source/target domain
U_s, V_s, U_t, V_t	source/target domain users and items
UID_s, VID_s	source domain unique identifiers of users and items
UID_t, VID_t	target domain unique identifiers of users and items
F_s, P_s, F_t, P_t	source/target domain profile features of users and items
G_s, G_t	source/target domain interest-item bipartite graph
M_s, M_t	adjacency matrices of G_s and G_t
G_v	cross-domain bipartite graph
M_v	adjacency matrix of G_v

With the development of deep learning, many researchers make use of deep learning technology to strengthen knowledge transfer. EMCDR[14] explicitly maps user representations from different domains via deep neural network. CoNet[6] proposes cross connections unit to enhancing the recommendation on both domains simultaneously. DARec[23] follows the idea of domain adaptation[16], extracting and transferring patterns from the rating matrix of related domains. PPGN[24] uses graph convolutional network[10] to explore the high-level connectivity between users and items on the joint interaction graph of two domains, and knowledge transfer by sharing user embedding. BiTGCF[12] propose a novel knowledge transfer module, which extends the flow of features from in-domain to inter-domain, and more importantly, considers the integration of uses' common features and domain-specific features. Compared with shallow cross-domain models, these deep learning-based methods usually show better performance due to stronger feature extraction capabilities.

3 PROBLEM DEFINITION

We begin with the definitions of terminologies. For clarity, the frequently used notations are summarized in Table 1. The first letter in upper case indicates a set of samples, and the first letter in lower case indicates a single sample.

Definition1. (Domain). Consider recommendation domain $D = \{U, V, E\}$, where U is the user set of D, V is the item set of D and E is the interactions between U and V. $U = \{UID, F\}$ is the set of users' unique identifier and long-term interests. $V = \{VID, P\}$ is the set of items' unique identifier and item profile features.

Definition2. (DSCR Problem). Given source domain $S = \{U_s, V_s, E_s\}$ and target domain $T = \{U_t, V_t, E_t\}$, where $|U_s \cap U_t| < 15\%|U_s \cup U_t|$, $V_s \cap V_t = \emptyset$, $||E_s||_1/|U_s| < 2$ and $||E_t||_1/|U_t| < 2$. The type and number of long-term interests of U_s and U_t are the same.

Definition3. (Goal of LGF Framework). Reducing the rating or ranking error in target domain T with the help of source domain S.

4 RESEARCH METHODS

In this section, we detail the proposed framework LGF. First, we present the Linking Domain. Subsequently we describe the Domain Mutual Information. Then we detail the Multi-task Generative Network. Next, we introduce the Adaptive Knowledge Transfer Network. Finally, we introduce the overall framework of LGF.

4.1 Linking Domain

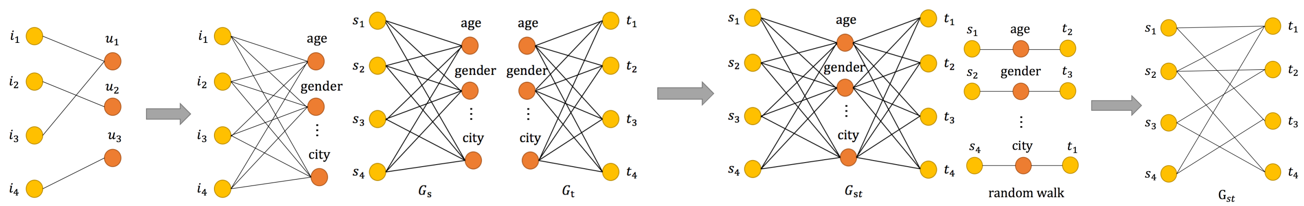
Many existing cross-domain recommendation methods adopt user IDs as a bridge linking two domains, as shown in Figure 1a. Abundant overlapped users help this method to establish a trusted domain association. However, In DSCR, there are very few overlapped users, and user behavior is sparse, which causes existing linking domain methods to fail to obtain a trusted domain association. Knowledge transfer based on untrusted domain associations has a greater risk and is likely to convey noisy information. As mentioned earlier, the behavior of users in DSCR is very sparse, which leads to their low tolerance to noise, and any heterogeneous information may cause negative transfer. In order to solve the above problems, we constructed a cross-domain bipartite graph to link the two domains based on the long-term and short-term interests of users. Compared with the method of linking domains based only on the user IDs, our method greatly enriches the links between the two domains and avoids the defects of the previous methods. The entire process of linking domain can be divided into three stages, namely graph transformation, graph fusion and graph pruning, as shown in Figure 3.

4.1.1 Graph Transformation. Overlapped users' behavior can be expressed as a user-item bipartite graph. The edges in the bipartite graph indicate that the user interacts with the item, and the weight of the edge indicates the number of interaction. We can transform user-item bipartite graph into a interests-item bipartite graph.

Specifically, if a user is connected to an item, we will connect the user profile features to the item. We apply this operation to the source and target domains, and get G_s and G_t.

4.1.2 Graph Fusion. There is a set of user profile features nodes in each of the two graphs, and the two sets of nodes are exactly the same. We merge the same two profile nodes into one node, and get G_{st}. User profile features are the bridge linking the two domains.

4.1.3 Graph Pruning. We adopt random walk in G_{st} to obtain the correlation between the source domain items and the target domain items. We take each source domain item as a starting point and perform multiple random walks with a walk length of 2 to obtain the target domain items associated with the source domain item. The first step adopts equation 1 as the transition probability to move from the source domain item to the user profile feature, and the second step adopts equation 2 as the transition probability to move from the user profile feature to the target domain item.



(a) Graph Transformation (b) Graph Fusion (c) Graph Pruning

Figure 3: Linking Domain (all graphs are weighted graphs)

According to the results of the random walk, we can construct a cross-domain bipartite graph G_v of the source domain items and the target domain items. If a source domain item reaches the target domain item N_a times in multiple random walks, connect them with an edge, and set the weight of the edge to N_a.

$$P(v_j|v_i) = \begin{cases} \dfrac{M_s^T ij}{\sum_{j \in N_{G_s}(v_i)} M_s^T ij} & , v_j \in N_{G_s}(v_i), \\ 0 & , v_j \notin N_{G_s}(v_i). \end{cases} \tag{1}$$

$$P(v_j|v_i) = \begin{cases} \dfrac{M_t^T ij}{\sum_{j \in N_{G_t}(v_i)} M_t^T ij} & , v_j \in N_{G_t}(v_i), \\ 0 & , v_j \notin N_{G_t}(v_i). \end{cases} \tag{2}$$

Where M_s and M_t are the adjacency matrices of G_s and G_t, $N_{G_s}(v_i)$ represents the set of neighbors of point v_i in G_s and $N_{G_t}(v_i)$ represents the set of neighbors of point v_i in G_t.

4.2 Domain Mutual Information

The cross-domain bipartite graph G_v depicts the correlation between two domains, which is a prerequisite for knowledge transfer between domains. We propose a method for measuring domain relevance based on G_v, called domain mutual information (DMI). Domain Mutual Information evaluates G_v from two perspectives of difference and diversity.

4.2.1 Difference of G_v. The composition of the source domain item information depends on the target domain items associated with it. We hope that the degree of association between each source domain item and different target domain items is different. If a source domain item has the same degree of association with all target domain items, the source domain item does not have any personalized information and cannot provide meaningful transfer information.

We use the weight on the edge of G_v to indicate the degree of association between the source domain item and the target domain item. We adopt variance to measure the difference between the degree of association between source domain items and different target domains items. We can evaluate the difference of G_v as follows:

$$Dist1 = log||(M_v - \frac{1}{|VID_s|} M_v H) \odot (M_v - \frac{1}{|VID_s|} M_v H)||_1 \tag{3}$$

Where M_v is the adjacency matrix of G_v, H is the matrix of ones and $\odot$ denotes the element-wise product of two vectors. VID_s is the set of source domain items, and $|VID_s|$ is the number of VID_s which is equal to the number of rows in M_v. The larger $Dist1$, the greater the difference in G_v.

4.2.2 Diversity of G_v. We hope that the most relevant target domain items of each source domain item have differences between them. Source domain items information is most affected by their most relevant target domain items. If the most relevant target domain items of each source domain item are the same, then the representation of the source domain item tends to be homogeneous and cannot provide meaningful information for the target domain. We can evaluate the diversity of G_v as follows:

$$Dist2 = \frac{|N_{G_v}^{K_1}(vids_1) \cup N_{G_v}^{K_1}(vids_2) ... \cup N_{G_v}^{K_1}(vids_{|VID_s|})|}{K_1 \times |VID_s|} \tag{4}$$

Where $N_{G_v}^{K_1}(vids_i)$ denotes the set of K_1 most relevant nodes among the neighbors of point $vids_i$ in G_v. The diversity of the G_v is positively correlated with $Dist2$.

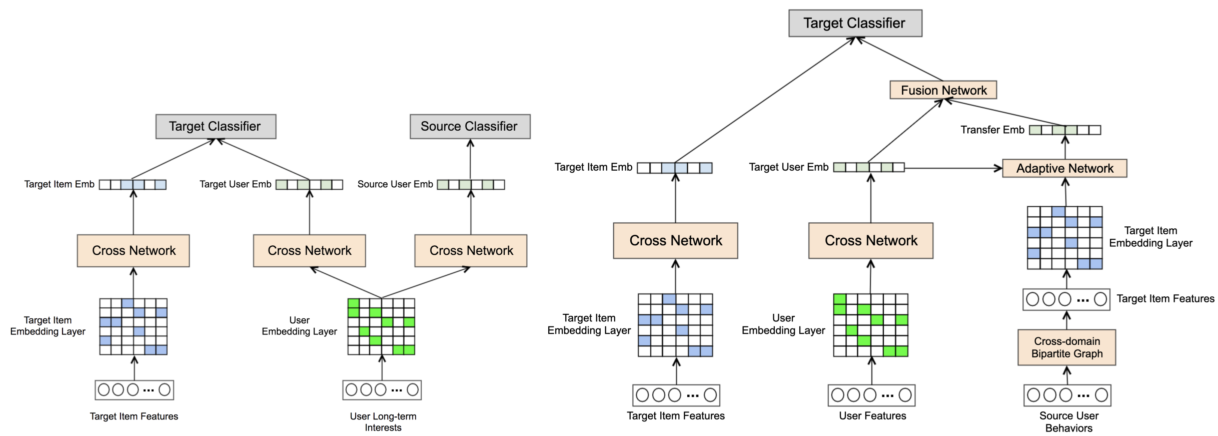
The calculation process of Domain Mutual Information is as follows:

$$DMI = Dist1 \times Dist2 \tag{5}$$

The value of DMI represents the degree of correlation between two domains. The greater the degree of correlation between the two domains, the easier it is to transfer knowledge. It is worth noting that in real industrial scenarios, $Dist1$ may be particularly large and $Dist2$ may be particularly small due to some hot items. If some target domain items are too hot, the most relevant target domain items of each the source domain item will be the same. At this time, you can filter out a certain number of hot items to obtain the real domain mutual information. Here we set K_3 as the ratio of filtering hot items.

4.3 Multi-task Generative Network

In the DSCR problem, a large part of the users in the target domain are non-overlapped users, and the existing source domain information can not provide them with useful information. We propose a multi-task generative network(MTG) to solve this problem. MTG consists of two parts, namely source domain multi-classification task and target domain recommendation task. Its structure is shown



(a) MTG (b) AKTN

Figure 4: LGF Framework (Item Fetures contains item_id and item profile features and User Fetures contains user_id and user profile features. Embedding layers with the same color in the same model share parameters)

in Figure 4a. The core idea of MTG is to use source domain multi-classification task to generate items that interact in the source domain for target domain non-overlapped users. For source domain multi-classification task, the training set is the data of source domain, and the test set is the data of target domain non-overlapped users. The distribution of the training set and the test set are inconsistent. Therefore, MTG adopts the target domain recommendation task as the regularization constraint of the source domain task to prevent the insufficient source domain task effect due to the domain adaptation[16] problem. The target domain task uses a part of the target domain data as training set, and shares the user embedding layer with the source domain task, which can prevent the source domain task from overfitting to the source domain data to a certain extent.

4.3.1 Source domain multi-classification task. The output of this task is a vector with the number of dimensions equal to the number of items in the source domain. Each dimension corresponds to a source domain item, and the value on the dimension corresponds to the probability of the user interacting with the item. We predict the probability of target domain non-overlapped users interacting with each source domain item, and set the K_2 items with the highest probability as the source domain behavior.

The overall calculation process is as follows:

$$X_s = E_U(fs) \tag{6}$$

$$q_s = M_S(X_s) \tag{7}$$

$$probs = C_S(q_s) \tag{8}$$

Where fs is user profile features, and M_S is the feature extractor. In this paper, we use cross network of DCN[21] as our feature extractor. C_S is a multi-class classifier.

The task is trained with the cross entropy in a supervised style:

$$\min_{E_U, M_S, C_S} L_1 = -\frac{1}{N_{ST}} \sum_{i=1}^{N_{ST}} \mathbf{y_i} \cdot log(probs_i) \tag{9}$$

where N_{ST} is the number of training data of MTG, $\mathbf{y_i}$ is the label of the i-th sample, which is a one-hot vector. $pred_i$ indicate the predicted value of the i-th item.

The purpose of this task is to generate items that interact in the source domain for target domain non-overlapped users. However, *user_id* is a unique identifier of a user and the *user_id* of non-overlapped users will not appear in the training data, so we do not use it for training.

It is worth noting that because the dimension of the task output vector is equal to the number of source domain items, when the number of source domain items is large, it may cause difficulty in task training and online deployment. We can set the dimension of the output vector to the number of categories of the source domain items to prevent the above problems.

4.3.2 Target Domain Recommendation Task. The overall calculation process is as follows:

$$X_u, X_v = E_U(ft), E_V(pt) \tag{10}$$

$$q_u, q_v = M_U(X_u), M_V(X_v) \tag{11}$$

$$q = concat([q_u^T, q_v^T])^T \tag{12}$$

$$pred = C_T(q) \tag{13}$$

Where ft is the user profile features, and pt is the item profile features. E_U and E_V are the embedding layers. M_U and M_V are the feature extractors used to obtain high-order nonlinear information q_u and q_v. C_T is a classifier used to predict ranking score (i.e.CTR).

The task is then trained with the standard log-loss in a supervised style:

$$\min_{E_U,E_V,M_U,M_V,C_T} L_2 = -\frac{1}{N_{ST}} \sum_{i=1}^{N_{ST}} y_i \cdot log(pred_i) + (1-y_i) \cdot log(1-pred_i) \tag{14}$$

where y_i and $pred_i$ indicate the label and predicted value of the i-th sample respectively.

These two tasks share the user embedding layer E_U. The training users of MTG are the source domain users, and the training data are the positive samples of the training users in the source domain and all the data of the training users in the target domain. For source domain non-overlapped users, we randomly sample the target domain items for him as the negative instances in target domain task. The number of negative samples is equal to the number of positive samples for the user in the source domain.

The overall loss of MTG is as follows:

$$L_{MTG} = L_1 + \lambda L_2 \tag{15}$$

Where λ is a hyper-parameter.

4.4 Adaptive Knowledge Transfer Network

The source domain information is a supplement to the target domain. We need to reduce the influence of redundant information similar to the target domain as much as possible, and extract the more complementary information of the target domain. Based on this idea, we propose an adaptive knowledge transfer network(AKTN) whose structure is shown in Figure 4b.

There are three key points: source item representation, adaptive source behavior representation, and knowledge transfer.

4.4.1 Source Item Representation. In DSCR, user behavior is extremely sparse, resulting in low tolerance for heterogeneous noise. Therefore, we require the source item representation to be composed of information in the target domain to ensure that no noise is mixed. The calculation of source item representation is as follows:

$$X_b = E_V(vt_b), \quad b \in N_{G_v}^{K_1}(a) \tag{16}$$

$$q_b = M_V(X_b), \quad b \in N_{G_v}^{K_1}(a) \tag{17}$$

$$q_a = \frac{\sum_{b\in N_{G_v}^{K_1}(a)} M_{vab} \times q_b}{\sum_{b\in N_{G_v}^{K_1}(a)} M_{vab}} \tag{18}$$

Where $N_{G_v}^{K_1}(a)$ denotes the set of K_1 most relevant nodes among the neighbors of item a in G_v, vt_b contains the unique identifier $vids_b$ and profile features of item b.

4.4.2 Adaptive Source Behavior Representation. Let $R_u = \{r_1, ..., r_{K_2}\}$ denote the set of K_2 items that user u recently interacted in the source domain. The essence of knowledge transfer is data fusion to make up for the lack of information in a single domain. Complementarity of data between domains can obtain more complete global information. The source behavior are the supplement to the target domain information. Different items in the source behavior supplement the information in the target domain to different degrees.

We propose adaptive network to achieve the above goals, the source behavior representation of user u is calculated as follows:

$$X_u = E_U(ut_u) \tag{19}$$

$$q_u = M_U(X_u) \tag{20}$$

$$\mathrm{z_i = V_1^T Tanh(W_1 q_u + U_1 q_{r_i})} \quad \forall \mathrm{r_i} \in \mathrm{R_u} \tag{21}$$

$$\mathrm{w_i} = \frac{\mathrm{e^{z_i}}}{\sum_{\mathrm{r_i \in R_u}} \mathrm{e^{z_i}}} \tag{22}$$

$$transfer_u = \sum_{i}^{K} w_i \times q_{r_i} \tag{23}$$

Where ut_u contains the unique identifier $uids_u$ and profile features of user u and q_u is the embedding of u. q_{r_i} denotes the embedding of r_i, w_i represents the degree to which q_{r_i} complements q_u and $transfer_u$ is the source behavior representation of u.

4.4.3 knowledge transfer. Next, we need to fuse the knowledge from the source domain and predict the user's interest. For a given pair of user u and item v, the target domain recommendation task fusing source domain information is calculated as follows:

$$weight = \sigma(W_2(concat([q_u^T, transfer_u^T)^T]) + b_2) \tag{24}$$

$$f_u = (1 - weight) \times q_u + weight \times transfer_u \tag{25}$$

$$q = concat([q_u^T, q_v^T])^T \tag{26}$$

$$score = C_T(q) \tag{27}$$

Where f_u the cross behavior representation of u.

The loss of AKTN is as follows:

$$\min_{E_U,E_V,M_U,M_V,V_1,W_1,U_1,W_2,b_2,C_T} L_{AKTN} = -\frac{1}{N_T} \sum_{i=1}^{N_T} y_i \cdot log(score_i) + (1-y_i) \cdot log(1-score_i) \tag{28}$$

4.5 LGF Framework

We now summarize our proposed LGF framework. The right of Figure 3c is the cross-domain bipartite graph G_v which shows the correlation of the source domain and target domain. The DMI of G_v represents the correlation between two domains. And based on the DMI, G_v can be further processed to describe the correlation between domains more accurately.

LGF Framework uses a two-stage training method which is shown in Figure 4 to first train MTG to generate user behavior in the source domain, and then train AKTN with the assistance of G_v to transfer the source domain knowledge to the target domain.

5 EXPERIMENTS

In this section, we evaluate the effect of LGF on the real dataset of DiDiChuxing, and deploy an online A/B test. We conduct experiments to answer the following research questions:

RQ1 How does LGF perform compared to the related methods?

RQ2 Is domain mutual information valuable to cross-domain recommendation?

RQ3 What effect does λ in MTG have on generating source domain behavior?

RQ4 Has LGF made any gains in the real scenario?

Table 2: target domain dataset

Users	Training set			Test set		
	percentage of users	percentage of data	average interactions	percentage of users	percentage of data	average interactions
Overlapped users	8%	12%	1.9415	13%	16%	0.0557
Source domain non-overlapped users	37%	23%	0	31%	22%	0.0152
Target domain non-overlapped users	55%	65%	1.7255	56%	62%	0.0579
All users	100%	100%	1.35	100%	100%	0.0482

Table 3: Offline performance AUC of compared methods

Users	average interactions	LGF-single	MV-DNN1	MV-DNN2	MLP++1	MLP++2	CoNet1	CoNet2	CNGAN	LGF
Overlapped users	0.0557	0.7326	0.7336	0.7311	0.7345	0.7323	0.7369	0.7376	0.7135	**0.7389(+0.86%)**
Source domain non-overlapped users	0.0152	0.7461	0.7232	0.7108	0.7035	0.7096	0.7028	0.7142	0.7236	**0.7514(+0.71%)**
Target domain non-overlapped users	0.0579	0.7131	0.702	0.6905	0.6785	0.6985	0.6817	0.7028	0.6771	**0.7238(+1.50%)**
All users	0.0482	0.7340	0.7288	0.7101	0.7137	0.7214	0.7155	0.7014	0.7122	**0.7433(+1.27%)**

5.1 Datasets

In order to verify the effectiveness of LGF, we choose the music recommendation (DiDi-Music) and the content recommendation (DiDi-Content) from DiDiChuxing. DiDi-Music and DiDi-Content are the recommendation scenarios that serve tens of millions of drivers. The source domain is DiDi-Music and the target domain is DiDi-Content. For the above recommendations, we separately extracted the online logs for 30 days, the first 25 days for the training set, and the remaining 5 days for the test set. Specifically, DiDi-Music dataset has 3 million interaction records and DiDi-Content dataset has 7 million interaction records.

The average number of user interactions in the source domain dataset is 1.79. The information of the target domain dataset is shown in Table 2. It is worth noting that the overlapped users in the test set of the target domain are overlapped users in the training set, rather than users who interact on the test sets on both domains. We take the items that the user has interacted with as positive instance, and the items that have been exposed but not interacted with as negative instance. The users in DiDi-Music and DiDi-Content have 11 profile features. The items in DiDi-Music and DiDi-Content have 11 and 0 profile features respectively.

The DSCR problem we proposed is very common and important in real industrial scenarios. However, no public dataset meets the definition of DSCR. In the future we will open our dataset[1].

5.2 Evaluation Metrics

We use AUC (area under ROC) and CTR as offline and online evaluation metric respectively.

5.3 Baseline

We compare single-domain, traditional cross-domain, and generative cross-domain methods. Traditional cross-domain refers to the methods of overlapped users, while generative cross-domain refers to the methods of generating source domain information for target domain non-overlapped users. Most of the existing cross-domain methods do not use profile features. In order to fairly compare the effects of baseline methods and LGF, we adopt the cross-domain methods based on deep learning, because we can integrate the profile features into these methods. It is worth noting that in this paper, the feature extractors of all models are cross network of DCN[21].

5.3.1 Single-domain method. We use the same model as the target domain recommendation task as our single-domain method, here we call it LGF-single.

(1) LGF-single: Its structure is the same as the target domain recommendation task.

5.3.2 Traditional cross-domain methods. All the following methods use the profile features of the user and the item.

(2) MV-DNN[1]: It extends the Deep Structured Semantic Model[8] and has a multi-tower matching structure.

(3) MLP++[6]. It combines two DNNs with shared user embeddings across domains.

(4) CoNet[6]. It proposes cross connections units based on cross-stitch networks[15], and adds cross connections units on MLP++ to enable dual knowledge transfer.

The above methods are suitable for overlapped users, but there are non-overlapped users in DSCR, so we need to do some processing on the training data. For non-overlapped users, we have two processing methods. One is to randomly select negative instances in another domain. The number of negative instances is equal to the number of samples of users in the training set. Another method is to specify an additional empty item as its negative instance. For a method M, denoted as M_1 and M_2 in these two data processing methods.

5.3.3 Generative cross-domain methods. In order to compare the effects of CnGAN and MTG and reduce other interference factors, we replaced the MTG with CnGAN in the LGF framework.

(5) CnGAN[18]: It uses GAN[5] to generate source domain preferences for target domain non-overlapped users.

5.4 Hyper-parameter Setting

We implement our model with Tensorflow[2]. To enable any one to reproduce the experimental results, we have attached all the hyper-parameters for each model in the table 4.

[1]https://github.com/mpt-dataset/LGF

[2]https://github.com/mpt-algorithm/LGF

Table 4: Hyper-parameter settings

Model	Hyper-parameter
General	embedding size=16 batchsize=4096 optimizer=Adam learning rate=0.001 feature extractor=cross network
NB-single	None
MV-DNN	None
MLP++	nets = [16, 16, 16] dropout=0.1
CoNet	Cross Connections Unit layer = 3 relationship matrices_L1=0.1
CnGAN	Generator=[16,16,16]
LGF	K1=2 K2=3 K3=2% λ = 0.6

5.5 Performance Evaluation (RQ1)

In this section, we evaluate the performance of LGF. The evaluation results of all the methods are presented in Table 3. From the results, we can see that most of traditional cross-domain methods have benefits on overlapped users. Among them, CoNet has achieved the best results, which also shows that the cross connections units can well integrate the knowledge of the two domains. However, traditional cross-domain methods perform poorly on non-overlapped users, which further illustrates the limitation of the application scope of traditional cross-domain methods. CnGAN performs the worst, which shows that the data of overlapped users is not enough to train a good generator.

LGF defeats all baselines on overlapped users and source domain non-overlapped users, which shows that the cross-domain bipartite graph is a good representation of the correlation between domains and provides strong support for knowledge transfer. LGF has made a big improvement in target domain non-overlapped users, which proves the correctness of the idea of generating users' behavior in the source domain. LGF has the smallest improvement on the source domain non-overlapped users. This is because the interactions of this class of users is the most sparse and the room for improvement is the smallest.

5.6 Impact of domain mutual information (RQ2)

In this section we analyze the value of domain mutual information to the cross-domain recommendation. DMI is an evaluation indicator that measures the amount of information in a cross-domain bipartite graph, which can guide the construction of a cross-domain bipartite graph.

DMI is affected by two hyper-parameters, namely K_2 and K_3. K_2 is a hyper-parameter for evaluating the diversity of information in G_v, and it is also the number of users behavior generated by MTG. K_3 is the proportion of filtered hot items in the target domain, which affects the difference and diversity of information in G_v. Here we fix other hyper-parameters and adjust the values of K_2 and K_3 to observe the relationship between DMI and AUC. The result is shown in Table 5.

Table 5: Relationship between DMI and AUC

K_3	$K_2 = 1$		$K_2 = 2$		$K_2 = 3$	
	DMI	AUC	DMI	AUC	DMI	AUC
0%	0.4053	0.7346	0.4581	0.7356	0.4072	0.7351
2%	0.7201	0.7373	0.6778	0.7359	**0.8472**	**0.7433**
4%	0.7778	0.7393	0.7389	0.7382	0.8213	0.7406
6%	0.6896	0.7359	0.7259	0.7367	0.6896	0.7351
8%	0.5541	0.7358	0.7959	0.7401	0.7959	0.7403
10%	0.773	0.7389	0.605	0.7354	0.8402	0.7406

It can be seen from the results that when the gap between DMIs is large, AUC and DMI are positively correlated. However, when the gap between DMIs is not large, the ordering between AUCs is not absolutely consistent with the ordering between DMIs. For example, when $K_3 = 2\%$ and $K_2 = 1$, DMI is 0.7201 and AUC is 0.7373, and when $K_3 = 6\%$, $K_2 = 2$, DMI is 0.7259 and AUC is 0.7367. This situation may be caused by random errors. The Spearman's correlation coefficient of DMI and AUC is 0.959.

In addition, when the DMI is small, the recommendation effect is also improved compared to LGF-single, which further illustrates the superiority of our framework. Based on the above analysis, DMI is positively correlated with cross-domain recommendation results, which can guide the selection of hyper-parameters to a certain extent.

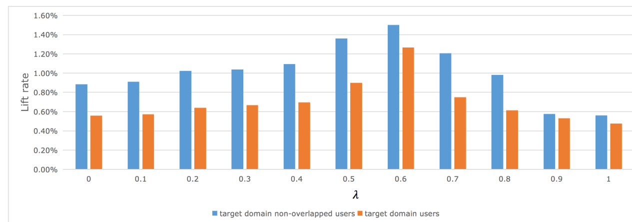


Figure 5: Comparison of test AUC by varying λ

5.7 Impact of λ (RQ3)

In MTG, the target domain recommendation task is used as the regularization constraint of the source domain multi-classification task to prevent the source domain multi-classification task from overfitting to the data of the source domain. This section evaluates the impact of λ on MTG. Since target domain non-overlapped users have no real source domain behavior and cannot directly evaluate the effect of MTG, we use the AUC lift rate of LGF for evaluation. The calculation is as follows:

$$\Delta lift = (AUC_{LGF} - AUC_{LGF-single})/AUC_{LGF-single} \qquad (29)$$

Where AUC_{LGF} means the test AUC value of LGF and $AUC_{LGF-single}$ means the test AUC value of LGF-single.

As shown in Figure 5, we set the keep λ from 0 to 1.0 with increments of 0.1 and it significantly affects the performance of lift rate of target domain non-overlapped users and target domain users.

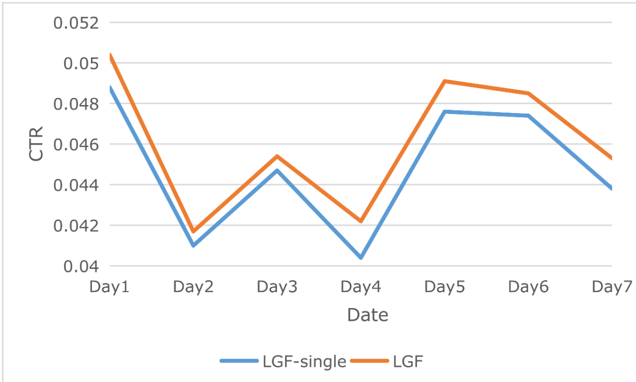


Figure 6: Online performance of compared methods

The lift rate of both of them first increase and then decrease with the increase of λ, and both achieve the best results at the λ of 0.6. When λ is 0, i.e., no constraints are imposed on the source domain multi-classification task, LGF cannot achieve good performance, which indicates that the source domain multi-classification task overfits the source domain data. When the λ tends to 1, MTG pays too much attention to the target domain recommendation task, which leads to underfitting of the source domain multi-classification task. The above results show that the target domain task in MTG can prevent overfitting of the source domain task.

5.8 Online Evaluation (RQ4)

To evaluate the real online performance of LGF, we perform online A/B tests on the DiDi-content. It can be seen from the offline experiments that LGF-single is the best baseline. Therefore, we use LGF-single as the baseline in the online test. The evaluation results are presented in Figure 6. LGF achieves the better CTR values in each day of the test week.

6 CONCLUSION

In this paper, we are the first to propose distant sparse cross-domain recommendation (DSCR). In DSCR, the proportion of overlapped users is small, and the data is very sparse, where the existing methods that cannot solve this problem. We further the linking and generation framework (LGF). LGF builds a cross-domain bipartite graph based on the user's long-term and short-term interests, linking the source and target domain. And we proposes domain mutual information (DMI) to measure the correlation between domains, which is the first proposed indicator to evaluate correlation between domains in the cross-domain recommendation field. In order to solve the problem of excessive proportion of non-overlapped users in DSCR, LGF proposes multi-task generation network (MTG) to generate source domain behaviors for non-overlapped users in the target domain. Finally, LGF uses adaptive knowledge transfer network (AKTN) to capture the beneficial information in the source domain.

REFERENCES

[1] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th international conference on world wide web*. 278–288.
[2] Castanedo Federico. 2013. A Review of Data Fusion Techniques. *Scientificworldjournal* 2013 (2013), 704504.
[3] Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, and Francesco Ricci. 2012. Cross-domain recommender systems: A survey of the state of the art. In *Spanish conference on information retrieval*. sn, 1–12.
[4] Wenjing Fu, Zhaohui Peng, Senzhang Wang, Yang Xu, and Jin Li. 2019. Deeply Fusing Reviews and Contents for Cold Start Users in Cross-Domain Recommendation Systems. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019), 94–101.
[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. *Advances in Neural Information Processing Systems* 3 (2014), 2672–2680.
[6] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. (2018).
[7] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. 2013. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd international conference on World Wide Web*. 595–606.
[8] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 2333–2338.
[9] Mohsen Jamali and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*. 135–142.
[10] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
[11] Tzu-Heng Lin, Chen Gao, and Yong Li. 2019. Cross: Cross-platform recommendation for social e-commerce. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 515–524.
[12] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 885–894.
[13] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*.
[14] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach.. In *IJCAI*. 2464–2470.
[15] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. 2016. Cross-stitch Networks for Multi-task Learning. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[16] Sinno Jialin Pan and Qiang Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.
[17] Dilruk Perera and Roger Zimmermann. 2017. Exploring the use of Time-Dependent Cross-Network Information for Personalized Recommendations. In *the 2017 ACM*. 1780–1788.
[18] Dilruk Perera and Roger Zimmermann. 2019. CnGAN: Generative Adversarial Networks for Cross-network User Preference Generation for Non-overlapped Users. In *The Web Conference 2019*.
[19] Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li. 2012. SocialTransfer: Cross-domain transfer learning from social streams for media applications. In *Acm International Conference on Multimedia*.
[20] A. P Singh and G. J Gordon. 2008. Relational learning via collective matrix factorization. In *Acm Sigkdd International Conference on Knowledge Discovery Data Mining*.
[21] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep Cross Network for Ad Click Predictions. In *ADKDD'17*.
[22] Xinghua Wang, Zhaohui Peng, Senzhang Wang, Philip S Yu, Wenjing Fu, and Xiaoguang Hong. 2018. Cross-Domain Recommendation for Cold-Start Users via Neighborhood Based Feature Mapping. (2018).
[23] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. DARec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns. *arXiv preprint arXiv:1905.10760* (2019).
[24] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-domain recommendation via preference propagation GraphNet. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2165–2168.
[25] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *SIGIR '20: The 43rd International ACM SIGIR conference on research and development in Information Retrieval*.