Cross-domain Recommendation Without Sharing User-relevant Data

Chen Gao
Beijing National Research Center for Information Science and Technology (BNRist),
Department of Electronic Engineering, Tsinghua University
gc16@mails.tsinghua.edu.cn

Xiangning Chen
Beijing National Research Center for Information Science and Technology (BNRist),
Department of Electronic Engineering, Tsinghua University

Fuli Feng
School of Computing, National University of Singapore

Kai Zhao
Beijing National Research Center for Information Science and Technology (BNRist),
Department of Electronic Engineering, Tsinghua University

Xiangnan He
School of Information Science and Technology, University of Science and Technology of China
xiangnanhe@gmail.com

Yong Li
Beijing National Research Center for Information Science and Technology (BNRist),
Department of Electronic Engineering, Tsinghua University
liyong07@tsinghua.edu.cn

Depeng Jin
Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University

ABSTRACT

Web systems that provide the same functionality usually share a certain amount of items. This makes it possible to combine data from different websites to improve recommendation quality, known as the *cross-domain recommendation* task. Despite many research efforts on this task, the main drawback is that they largely assume the data of different systems can be **fully shared**. Such an assumption is unrealistic — different systems are typically operated by different companies, and it may violate business privacy policy to directly share user behavior data since it is highly sensitive.

In this work, we consider a more practical scenario to perform cross-domain recommendation. To avoid the leak of user privacy during the data sharing process, we consider sharing only the information of the item side, rather than user behavior data. Specifically, we transfer the item embeddings across domains, making it easier for two companies to reach a consensus (e.g., legal policy) on data sharing since the data to be shared is user-irrelevant and has no explicit semantics. To distill useful signals from transferred item embeddings, we rely on the strong representation power of neural networks and develop a new method named as NATR (short for ***N****eural* ***A****ttentive* ***T****ransfer* ***R****ecommendation*). We perform extensive experiments on two real-world datasets, demonstrating that NATR achieves similar or even better performance than traditional cross-domain recommendation methods that directly share user-relevant data. Further insights are provided on the efficacy of NATR in using the transferred item embeddings to alleviate the data sparsity issue.

CCS CONCEPTS

• **Information systems → Collaborative filtering**; **Recommender systems**; • **Security and privacy → Privacy-preserving protocols**.

KEYWORDS

Cross-domain Recommendation, Privacy Preserving, Deep Learning

ACM Reference Format:
Chen Gao, Xiangning Chen, Fuli Feng, Kai Zhao, Xiangnan He, Yong Li, and Depeng Jin. 2019. Cross-domain Recommendation Without Sharing User-relevant Data. In *Proceedings of the 2019 World Wide Web Conference (WWW'19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3308558.3313538

1 INTRODUCTION

In the current Web ecosystem, it is common that some websites have a certain degree of homogeneity in their functionality and provided information. For example, there are many overlapped hotels on Trip.com and Booking.com, overlapped movies on IMDb and Douban[1], and overlapped products on Amazon and eBay. From the perspective of building recommendation services, it means

[1]Douban.com is a Chinese website that allows registered users to rate movies, music, books, etc.

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313538

that the models for such two homogeneous domains are dealing with many items that are the same. This provides opportunities to improve the recommendation quality by enriching data. For example, if domain A does not have sufficient data on some items (*i.e.*, sparse or cold-start items) while the other domain B does have, e.g., a movie first released in US may have many ratings on IMDb but not on Douban, then the recommendation for these items on domain A can be potentially improved by integrating the data of domain B. The task of leveraging auxiliary data from other domains to improve recommendation quality of the target domain is known as the *cross-domain recommendation* [1, 10, 15, 32, 35, 43].

Existing works on cross-domain recommendation have primarily focused on directly aggregating data from multiple domains [1, 15, 32, 35, 43]. In other words, these methods assume that during model training of the target domain, user behavior data of other domains are directly accessible. For example, the representative Collective Matrix Factorization (CMF) [32] method extends Matrix Factorization (MF) by jointly learning user embeddings and item embeddings from the user-item interaction matrix of multiple domains. Despite effectiveness, the assumption that user behavior data can be fully shared across domains is questionable. Typically, different domains (websites) are operated by different companies, and thus it is difficult to let them share user behavior data due to the constraint of company policy.

In this work, we aim to provide a more realistic solution for cross-domain recommendation. To avoid any chance of leaking user privacy, we abandon the sharing of user-relevant data, neither behavior logs nor demographic attributes. However, this will pose challenges to transfer the collaborative filtering (CF) signal from one domain to another, since CF is typically modeled through mining user-item interaction data (e.g., user purchase and click logs). To address this technical challenge, we propose to share the item embeddings, which are learned by reconstructing user-item interaction matrix. The advantages are two-fold: 1) item embeddings can still encode certain CF signal by reflecting item similarities based on user behaviors (e.g., which items are frequently co-rated by users)[2], and 2) item embeddings are latent vectors that have no explicit semantics; as such, the risk of leaking user privacy can be kept to a minimum, which makes it easier for two companies to reach a legal policy for data sharing[3].

Our proposed solution, which has three steps, is illustrated in Figure 1. In the first step, an embedding-based recommender model, MF for example, is trained on the user-item interaction matrix of the auxiliary domain to obtain item embeddings. In the second step, item embeddings of the auxiliary domain are sent to the target domain; note that only the embeddings of overlapped items are necessary to be sent, which are subjected to the data-sharing policy between two companies. Finally, the target domain trains a recommender model with the consideration of the transferred item embeddings. The first two steps are straightforward to implement, and the main challenges lie in how to design a model to effectively incorporate transferred item embeddings in the last step. We summarize the key challenges as follows.

[2] A representative example of using the signal in item embeddings for recommendation is the item-based CF methods [13, 17].

[3] https://www.nytimes.com/interactive/2018/06/03/technology/facebook-device-partners-users-friends-data.html

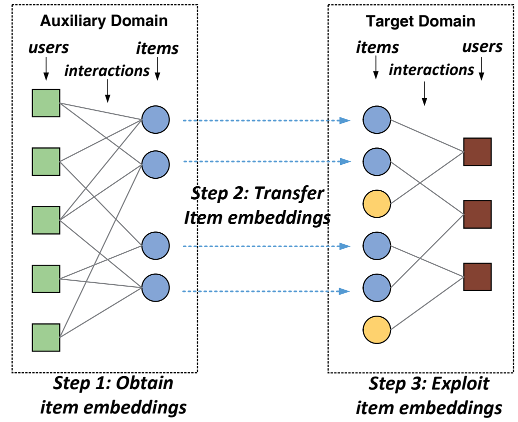


Figure 1: Illustration of our solution for cross-domain recommendation without sharing user-relevant data.

- **Unclear predictive signal of transferred item embeddings**. It is unclear whether and which transferred item embeddings contain useful signal in estimating a user's preference on an item in the target domain. Note that one motivation of conducting cross-domain recommendation is to alleviate the data sparsity problem in the target domain. However, the data sparsity problem may also exist in the auxiliary domain for some items, or the other way round the data in the target domain is already sufficient and does not require extra supplement. As such, it is challenging to distill useful signal from the transferred item embeddings and integrate them into the predictive model of the target domain.
- **Varying importance of transferred item embeddings**. As mentioned, the data of the auxiliary domain is not oracle — it is likely that user behaviors on some items are sparse and are insufficient to learn good embeddings for them. As such, it is a common case that the quality of item embeddings varies, where items of many users behaviors may have good quality and vice versa. Since it is already difficult to judge the quality of learned item embeddings for the auxiliary domain, it becomes even more challenging for the target domain to utilize such unknown- and varied- quality item embeddings well.
- **Embedding dimension discrepancy in latent space**. The data for training in the two domains may be of different scale and have different distribution. Therefore, the optimal embedding size for the two domains may be different. As such, existing cross-domain recommendation solutions that perform regularization on embedding matrices will fail [26]. Moreover, even though we restrict their embedding sizes to be the same, the semantics of their embedding dimensions are different and cannot be directly aligned.

To solve the above mentioned challenges, we design a novel model named **N**eural **A**ttentive-**T**ransfer **R**ecommendation (NATR). Briefly, our proposed method relies on the strong representation power of neural networks and discriminative power of attention mechanisms to leverage the transferred item embeddings. Specifically, we design 1) a domain-level attention unit to dynamically adjust the importance of the predictive signal of the two domains; 2) an item-level attention unit to determine which embeddings of transferred items are more useful in constructing user representation for further prediction; and; 3) a domain adaption layer to bridge the discrepancy between the embedding space of the two domains. By tailoring our solution for addressing the highlighted challenges, our NATR method demonstrates its strong performance in cross-domain recommendation, and meanwhile preserves user privacy during data sharing.

To summarize, the main contributions of this work are as follows.

- We present a new paradigm for cross-domain recommendation without sharing user-relevant data, in which only item-side data can be shared across domains. To allow the transferring of CF signal, we propose to share the item embeddings which are learned from user-item interactions of the auxiliary domain.
- We propose a new solution NATR to resolve the key challenges in leveraging transferred item embeddings. The two-level attention design allows NATR to distill useful signal from transferred item embeddings, and appropriately combine them with the data of the target domain.
- We conduct extensive experiments on two real-world datasets to demonstrate our proposed method. More ablation studies verify the efficacy of our designed components, and the utility of transferred item embeddings in addressing the data sparsity issue.

The remainder of this paper is as follows. We first formulate the research problem in Section 2. We then elaborate our proposed method in Section 3. We conduct experiments in Section 4, before discussing related work in Section 5. Lastly, we conclude this paper in Section 6.

2 PROBLEM FORMULATION

We first introduce some notations used in the paper. We represent matrices, vectors, and scalars as bold capital letters (*e.g.,* $\mathbf{X}$), bold lower letters (*e.g.,* $\mathbf{x}$), and normal lowercase letters (*e.g.,* x), respectively. If not otherwise specified, all vectors are in a column form; $\mathbf{X}^T$ denotes the transpose of $\mathbf{X}$. We use symbols σ, $ReLU$, and $\odot$ to denote the *sigmoid* function, rectifier function, and element-wise production operation, respectively.

2.1 Cross-domain Recommendation

A typical problem setting of *cross-domain recommendation* is leveraging the data from an *auxiliary* domain to facilitate the recommendation quality in a *target* domain with overlapped items. In the *target domain*, where M and N denote the number of users and items, respectively, we have a user-item interaction matrix $\mathbf{Y}^t \in \mathbb{R}^{M \times N}$ with a binary value at each entry defined as,

$$y_{ui}^t = \begin{cases} 1, & \text{if } u \text{ has interacted with } i; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Similarly, in the *auxiliary domain*, we have another binary user-item interaction matrix $\mathbf{Y}^a \in \mathbb{R}^{K \times L}$, where K and L are the number of users and items. Note that a portion of L items also occurs in the target domain, which are named as *bridge items*. From the interaction matrices $\mathbf{Y}^t$ and $\mathbf{Y}^a$, the goal of cross-domain recommendation is to learn a predictive function to estimate the likelihood that a given user u will interact with item i in the target domain.

2.2 Cross-domain Recommendation Without Sharing User-relevant Data

Distinct from the typical problem settings of cross-domain recommendation, we abandon the direct sharing of user behavior data (the user-item interaction matrix $\mathbf{Y}^a$ in the auxiliary domain). This is because directly sharing user behavior data may violate the business privacy policy of different companies operating the auxiliary and target domains. Instead, we propose a solution that only transfers the embeddings of bridge items which are offline learned in the auxiliary domain, as illustrated in Figure 1. We define the transferred item embedding matrix $\mathbf{Q}^a = [\mathbf{q}_1^a, \cdots, \mathbf{q}_N^a] \in \mathbb{R}^{D' \times N}$ as,

$$\mathbf{q}_i^a = \begin{cases} \tilde{\mathbf{q}}_i^a, & \text{if item } i \text{ is a bridge item;} \\ \mathbf{0}, & \text{otherwise;} \end{cases} \tag{2}$$

where $\mathbf{0} \in \mathbb{R}^{D'}$ is an all-zero vector and $\tilde{\mathbf{q}}_i^a \in \mathbb{R}^{D'}$ is the offline learned embedding of item i in the auxiliary domain. It should be noted that we organize the transferred item embeddings in $\mathbf{Q}^a$ in the same order of item IDs in the target domain to enable looking up an item embedding with its ID. Here we assume the availability of $\tilde{\mathbf{q}}_i^a$, *i.e.,* the company operating the auxiliary domain has employed an embedding-based recommendation system [19]. Note that the assumption is practical since embedding-based recommendation solutions are widely applied in the industry [4, 5].

After introducing the transferred item embeddings $\mathbf{Q}^a$ from the auxiliary domain, we formulate the problem of cross-domain recommendation without sharing user-relevant data as follows,

Input: The user-item interaction data in the target domain $\mathbf{Y}^t$, and the transferred item embeddings $\mathbf{Q}^a$ from the auxiliary domain.

Output: A predictive model to estimate the likelihood that a user u will interact with an item i in the target domain. Specifically, taking u, i, and $\mathbf{y}_u^t$ which is the interaction history[4] of u in the target domain, as input, the model has to predict,

$$\hat{y}_{ui}^t = f(u, i, \mathbf{y}_u^t), \tag{3}$$

where $\hat{y}_{ui}^t \in [0, 1]$ denotes the probability of interaction between user u and item i.

After obtaining the predictive model, we can use it to score all items for a given user u, and select the top-ranked (*i.e.,* with higher interaction probability) items as the recommendation results for u. It should be noted that there indeed exist user and item attributes in both the auxiliary and target domains. However, to simplify the scenario of the cross-domain recommendation task, we only emphasize the user-item interactions, which is a common setting of existing works [26, 43].

3 PROPOSED METHOD

To solve the problem of cross-domain recommendation without sharing user-relevant data, we rely on the strong representation ability of neural networks and devise a new solution, named NATR,

[4] Note that $\mathbf{y}_u^t$ is the transpose of the u-th row of $\mathbf{Q}^t$.

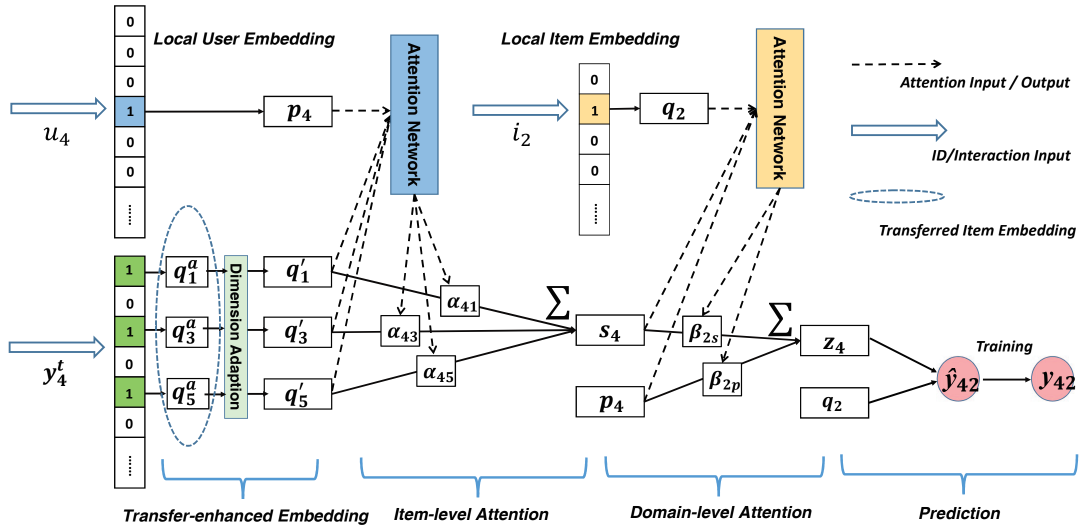


Figure 2: The architecture of our proposed Neural Attentive Transfer Recommendation model. (Here we take the prediction of user-item pair $(4, 2)$ as an example; user 4 has interacted with item 1, 3, and 5 in the target domain; then we utilize $\mathbf{q}_1^a$, $\mathbf{q}_3^a$, and $\mathbf{q}_5^a$ transferred from the auxiliary domain to assist predicting y_{42} in the target domain. Note that $\mathbf{q}^a$ is the only available auxiliary data in the target domain.)

exploiting the transferred item embeddings. Figure 2 illustrates the architecture of our proposed NATR model, which are made up of the following four layers.

- **Transfer-enhanced Embedding Layer**. We project sparse user and item representations into dense vectors. A dimension-adaption module is adopted to solve the dimension discrepancy problem of transferred item embeddings.
- **Item-level Attentive Layer**. To enrich user representations, we fuse the transferred embedding of items a user interacted with to an *additional user embedding* with an item-level attention unit to model the varying importance of items.
- **Domain-level Attentive Layer**. With a consideration of the diversity across domains, we make use of a domain-level attention unit to control the influence of predictive signals from two domains.
- **Prediction Layer**. Finally, we utilize an inner-product model as the predictive function since our work mainly focuses on devising a framework to exploit the transferred item embeddings.

In the following, we elaborate the details of the aforementioned four layers.

3.1 Transfer-enhanced Embedding Layer

Latent factor model (LFM) is one kind of general framework in collaborative recommender systems, which associates each user and item with real-valued vectors . Considering that LFMs have achieved success in a wide range of recommendation tasks [16, 17, 19, 24, 28], we project sparse user and item representations into real-valued vectors. Specifically, we first encode user ID (u) and item ID (i) into one-hot encodings as follows,

$$\mathbf{v}_u^U = \text{one-hot}(u), \mathbf{v}_i^I = \text{one-hot}(i), \tag{4}$$

where $\mathbf{v}_u^U$ ($\mathbf{v}_i^I$) $\in \mathbb{R}^N$ is a vector with all zero values except the u-th (i-th) entry with value 1. We then project the sparse one-hot encodings ($\mathbf{v}_u^U$ and $\mathbf{v}_i^I$) and multi-hot interaction history ($\mathbf{y}_u^t$) to *local embeddings* and *transferred embeddings*, respectively.

Local embeddings. To project the one-hot user (item) encoding, we employ an embedding layer, which is defined as a fully connected layer without bias term as follows,

$$\mathbf{p}_u = \mathbf{P}^T \mathbf{v}_u^U, \quad \mathbf{q}_i = \mathbf{Q}^T \mathbf{v}_i^I, \tag{5}$$

where $\mathbf{P} \in \mathbb{R}^{N \times D}$ and $\mathbf{Q} \in \mathbb{R}^{M \times D}$ are the parameters to be learned. The obtained embeddings $\mathbf{p}_u$ and $\mathbf{q}_i \in \mathbb{R}^D$ are named as *local embeddings* since they are learned merely with information from the target domain.

Transferred embeddings. In our problem, transferred embeddings of bridge items are the only auxiliary data accessible in the target domain. From the perspective of representative learning, there are two manners to leverage these item embeddings in collaborative filtering: user-based [24, 28] and item-based [13, 17]. Specifically, when predicting the probability that user u will choose item i in the target domain, user-based CF means directly combining embeddings of i of two domains to match $\mathbf{p}_u$ while item-based manner means matching $\mathbf{q}_i$ with transferred embedding of user's historically interacted items. There are two key aspects to make item-based manner a more convincing choice. First, when item i is not a bridge item, user-based CF solution cannot bring any help to prediction. Second, user-based CF can only distill implicit preferences while neglecting the explicit preferences of a user (*i.e.*, the historically interacted items), while item-based CF can extract explicit preferences through leveraging transferred embeddings of bridge items and multi-hot encoding of historical interactions to enrich user representation.

Therefore, we look up the transferred embedding $\mathbf{q}_j^a$ from $\mathbf{Q}^a$ for each item j with $y_{uj}^t = 1$. Note that $\mathbf{q}_j^a$ will be an all-zero vector if item j is not a bridge item (see Equation 2 for details). As mentioned above, there may exist the challenge of *embedding dimension discrepancy* across two domains, *i.e.,* $D \neq D^{'}$, since the training data in two domains may be of different scale and have different distribution. To solve this problem, we employ a fully-connected layer to adapt the dimension of transferred embeddings, which is formulated as follows,

$$\mathbf{q}_j^{'} = \mathbf{W}_0^T \mathbf{q}_j^a + \mathbf{b}_0, \tag{6}$$

where $\mathbf{W}_0 \in \mathbb{R}^{D^{'} \times D}$ and $\mathbf{b}_0 \in \mathbb{R}^D$ are learnable parameters of mapping matrix and bias. Note that $\mathbf{W}_0$ and $\mathbf{b}_0$ are the only parameters here as the transferred embedding matrix $\mathbf{Q}^a$ is learned offline in the auxiliary domain.

With the above neural components, we project the sparse one-hot and multi-hot encodings into local user and item embeddings, $\mathbf{p}_u$ and $\mathbf{q}_i$, and transferred item embeddings, $\{\mathbf{q}_j^{'} | y_{uj}^t = 1\}$. We introduce our item-based CF solution of leveraging transferred item embeddings detailedly in the following parts.

3.2 Item-level Attentive Layer

The key objective of embedding-based recommendation model is to capture relation between of user and item in the latent space [41], therefore it is critical to explicitly build the relation between transferred item embeddings and local user embedding in our problem. As mentioned above, motivated by item-based CF [13, 17], which encodes the historical interaction behaviors of a user to enrich the user representation, we fuse the transferred item embeddings $\{\mathbf{q}_j^{'} | y_{uj}^t = 1\}$ into an *additional user embedding* $\mathbf{s}_u$. Besides historical interactions, $\mathbf{s}_u$ also contains CF signals transferred from the target domain, which can further enhance user representation. Our first inspiration to calculate $\mathbf{s}_u$ is average pooling, a widely used modeling component in neural networks, formulated as follows,

$$\mathbf{s}_u = \frac{\sum_{\{j | y_{uj}^t = 1\}} \mathbf{q}_j^{'}}{|\mathbf{y}_u^t|}, \tag{7}$$

where $|\mathbf{y}_u^t|$ is the l_1-norm of vector $\mathbf{y}_u^t$, which equals to the number of items user u has interacted with.

However, as mentioned above, there exists another key challenge of *varying importance of embeddings.* Considering that different interacted items have embeddings with varying quality and varying importance to represent the preference of a given user, such naive operation may not work well in real scenario. Therefore, we apply a non-uniform coefficient when fusing the transferred item embeddings:

$$\mathbf{s}_u = \frac{\sum_{\{j | y_{uj}^t = 1\}} \alpha_{uj} \mathbf{q}_j^{'}}{|\mathbf{y}_u^t|}. \tag{8}$$

To model the various item importance in a user-sensitive fashion, here we introduce attention mechanism, which has achieved great success in recommendation tasks [3, 13, 40] and natural language processing [20, 37]. Specifically, the item-level attention unit learns a specific weight α_{uj} for every transferred item embedding $\mathbf{q}_j^{'}$ according to the following formulation,

$$\begin{aligned} \alpha_{uj} &= \frac{e^{a_{uj}}}{\sum_{\{k | y_{uk}^t = 1\}} e^{a_{uk}}}, \\ a_{uj} &= \mathbf{w}_1^T ReLU(\mathbf{p}_u \odot \mathbf{q}_j^{'}) + b_1, \end{aligned} \tag{9}$$

where $\mathbf{w}_1 \in \mathbf{R}^D$ and b_1 denote the weight matrix and bias of a fully connected layer. The input of the item-level attention unit is the interaction between the user and target item, which makes the learned attention score sensitive to the given user.

3.3 Domain-level Attentive Layer

After obtaining the *local* and *additional* user embeddings, our task become learning a prediction function based on these three embeddings. Different from traditional recommendation models, here we have an extra user embedding. To exploit two user embeddings, we fuse them into an *unified embedding.* This is inspired by some cross-domain recommendation models [33, 43], which have demonstrated that fusing embedding vector learned from multi-modal data is a simple but yet effective way to combine signals. Another option is to separately estimate the interaction probability with the two embeddings and fuse the predictions (late fusion). Here we employ early fusion that merges embeddings, allowing us to explicitly capture the interaction between two embeddings.

The *unified embedding* via fusion can be denoted as,

$$\mathbf{z}_u = \beta_{si}\mathbf{s}_u + \beta_{pi}\mathbf{p}_u, \ s.t., \ \beta_s + \beta_p = 1, \tag{10}$$

where β_{si} and β_{pi} are learnable weights for $\mathbf{s}_u$ and $\mathbf{p}_u$, respectively. The aim of β_{si} and β_{pi} is to balance the information from auxiliary and target domain regarding the target item i. In other words, these two weights are item-sensitive. We devise such design to address the key challenge of *unclear predictive signal* which has been mentioned before. That is, in real scenario evaluating different items needs varying amount of auxiliary information. Although we can directly learn β_{si} and β_{pi} by optimizing a final objective function, we rely on the promising representation ability of attention network to model them. Formally, a domain-level attention module is designed as,

$$\begin{aligned} \beta_{si} &= \frac{e^{b_{si}}}{e^{b_{si}} + e^{b_{pi}}}, \ \beta_{pi} = \frac{e^{b_{pi}}}{e^{b_{si}} + e^{b_{pi}}}, \\ b_{si} &= \mathbf{w}_2 ReLU(\mathbf{s}_u \odot \mathbf{q}_i) + b_2, \\ b_{pi} &= \mathbf{w}_2 ReLU(\mathbf{p}_u \odot \mathbf{q}_i) + b_2, \end{aligned} \tag{11}$$

where $\mathbf{w}_2 \in \mathbb{R}^D$ and b_2 are the parameters of the attention network. Note that the input of the attention network is the interaction between user ($\mathbf{s}_u/\mathbf{p}_u$) and item embeddings, which enables the learned attention scores to be sensitive to item i.

3.4 Prediction Layer

After the operation in aforementioned layer, our problem further turns to predict user interaction with two embeddings: unified user embedding vector $\mathbf{z}_u$ and item embedding vector $\mathbf{q}_i$. Here we adopt a predictive function to estimate y_{ui}^t which is the interaction probability between a given pair of user and item (u, i). Since our work mainly focuses on a general framework for cross-domain recommendation without sharing user-relevant data, we adopt a simple but widely-used inner product model, to estimate the value of y^t_{ui}, which is formulated as

$$\hat{y}^t_{ui} = \sigma(\mathbf{z}_u{}^T \mathbf{q}_i), \tag{12}$$

where σ is the *sigmoid* function. Note that this predictive function can be easily extended to more complicated ones, such as the multi-layer perceptron in [14].

To conclude, with three specially devised layers and a prediction layer, aforementioned three key challenges of our problem are addressed one by one.

3.5 Training

Objective Function. Following the probabilistic optimization framework [14, 28], we first define the likelihood function for an implicit interaction as follows,

$$P_r = \prod_{(u,i)\in\mathcal{Y}^t_+} \hat{y}^t_{ui} \prod_{(u,i)\in\mathcal{Y}^t_-} (1 - \hat{y}^t_{ui}), \tag{13}$$

where $\mathcal{Y}^+_t$ denotes the set of observed interactions in interaction matrix of the target domain $\mathbf{Y}^t$ (entries with value of 1), and $\mathcal{Y}^t_-$ denotes negative instances sampled from the unobserved interactions in $\mathbf{Y}^t$ (entries with value of 0). We further take the negative logarithm of the joint probability, and obtain the loss function (*a.k.a, logloss* [29]), which is widely used to optimize recommendation systems with implicit feedbacks [14, 17, 24], to be minimized as follows,

$$\mathcal{L} = -\Big(\sum_{(u,i)\in\mathcal{Y}^t_+} \log \hat{y}^t_{ui} + \sum_{(u,i)\in\mathcal{Y}^t_-} \log(1 - \hat{y}^t_{ui})\Big). \tag{14}$$

To prevent over-fitting, we adopt l_2 regularization on the parameters in the proposed neural model and obtain the overall objective function,

$$\Gamma = \mathcal{L} + \lambda \sum_{\Theta\in\{\mathbf{P},\mathbf{Q},\mathbf{W}_0,\mathbf{b}_0,\mathbf{w}_1,b_1,\mathbf{w}_2,b_2\}} \|\Theta\|^2_F. \tag{15}$$

Mini-batch Training. We adopt stochastic gradient descent (SGD) [34], a widely generic solver for neural models, to optimize our proposed NATR model in the mini-batch mode. To construct a mini-batch, we first sample a batch of historical user-item interaction pairs (u, i). For each (u, i), we then adopt a negative sampling technique [24], which is widely used to handle implicit feedbacks in existing researches [14, 28, 43], to randomly select unobserved items $\{i^{'}_1, \cdots i^{'}_n\}$ for user u with a sampling ratio of n. After the sampling, we obtain n triplets $\{(u, i, i^{'}_1), \cdots, (u, i, i^{'}_n)\}$ for each instance in the batch. With the constructed mini-batch, we take a gradient step to minimize the objective function.

4 EXPERIMENTS

In this section, we conduct extensive experiments on two real-world datasets to answer the following research questions:

- **RQ1:** How does our proposed NATR model perform compared with the state-of-the-art methods for cross-domain recommendation tasks?
- **RQ2:** Can the proposed NATR alleviate the data sparsity problem in the target domain?
- **RQ3:** What are the effects of the item-level and domain-level attention models in our proposed NATR?

In what follows, we first describe the experimental settings, and then answer the above three research questions.

Table 1: Statistics of our evaluation datasets.

Dataset	Item#	Auxiliary Domain		Target Domain	
		User#	Rec#	User#	Rec#
ML-NF	5,568	31,038	2,269,179	14,630	152,206
TC-IQI	4,851	35,398	314,621	19,999	78,429

4.1 Experimental Settings

4.1.1 Datasets. We experiment with two real-world datasets that both contain implicit interactions from two domains.

- **ML-NF Dataset.** MovieLens and Netflix are two popular platforms with movie recommendation services, in which there are a large portion of overlapped movies. Here we take MovieLens (ML) and Netflix (NF) as the *auxiliary* and *target* domains (*i.e.,* our target is to improve the recommendation performance in NF), respectively. We obtain user-movie interactions in ML and NF from two widely used public movie rating datasets[5][6]. Note that we identify movies with the same name in the two datasets as *bridge items*. Here we conduct whole-string matches to avoid wrong matches as possible. By filtering bridge items and their associated ratings[7], we reserve 5,568 movies, 14,630 ML users and 31,038 NF users in this dataset. Finally, we intentionally transform the rating data into binary (1/0 indicate whether a user has interacted with an item or not) to fit the problem setting of implicit feedback [16].
- **TC-IQI Dataset.** This dataset is collected by [43] to evaluate *cross-domain recommendation* performance of online video contents. In this dataset, there are historical interactions between users and videos from two mainstream video websites, iQiyi[8] (IQI) and Tencent Video[9] (TC) in China. To investigate the performance of facilitating recommendation performance in target domain via leveraging information from auxiliary domain, we regard *IQI* (*TC*) as the *target* (*auxiliary*) domain since interactions in IQI are sparser. Similarly, we filter *bridge items* via exact name matching across videos from these two domains, and only reserve bridge items and interactions associated with them. Note that videos in this dataset are professional production content (PGC) widely available on multiple websites.

After the above pre-processing steps, we obtain two final datasets for performance evaluation, the statistics of which are summarized in Table 1.

[5]https://grouplens.org/datasets/movielens/
[6]https://www.kaggle.com/laowingkin/netflix-movie-recommendation/data
[7]To better evaluate the performance, we follow the setting of previous works [27, 32] and only reserve the bridge items (*i.e.,* neglect items occurring only in one domain). However, our model is also suitable to perform recommendation for those items not overlapped, which has been introduced in detail in Section 3.1
[8]https://www.iqiyi.com
[9]https://v.qq.com

4.1.2 Evaluation Protocols. Following existing works [12, 31], we employ the widely used *leave-one-out* evaluation protocol in the evaluation stage. Similar with [14, 18], given a user in the target domain, we randomly sample 99 items that are not interacted with the user, each method ranks one test item among the 99 sampled items. We then adopt two metrics, *HR* and *NDCG*, which are widely used in the literature of recommendation [6, 14, 39], to report the ranking performance of each method:

- **HR@K:** *Hit Ratio* (HR) measures whether the test item is contained by the top-K item ranking list (1 for yes and 0 for no).
- **NDCG@K:** *Normalized Discounted Cumulative Gain* (NDCG) extends HR by assigning higher scores to the hits at higher positions in the ranking list.

It should be noted that we calculate HR@K and NDCG@K for each test user, and report the average ones over the whole user set.

4.2 Performance Comparison (RQ1)

4.2.1 Baselines. We compare the performance of our proposed NATR with five baselines, which can be divided into two groups: *single-domain* and *cross-domain*. Here *single-domain* methods refer to those which are merely trained with data from the target domain, while *cross-domain* methods jointly consider the data from both the target and auxiliary domains.

The compared *single-domain* methods are introduced as follows:

- **PMF [24].** *Probabilistic Matrix Factorization* (PMF) is a MF based model which exploits negative sampling to handle implicit interaction data. It adopts *logloss* as the loss function and samples several negative items with a ratio when a positive item is fed for training. We tune the learning rate and regularizer and report the best testing performance.
- **GMF.** *Generalized Matrix Factorization* (GMF) is one of the variants of NCF (Neural Collaborative Filtering) [14], which is the state-of-the-art solution for recommendation tasks with implicit feedbacks. This method assign various weights for different dimensions in the dot-product prediction function, which can be regarded as a generalization of vanilla MF. We optimize this model and tune its associated hyper-parameters similarly with the paper.
- **NATR-local.** As mentioned in Section 3.1, our NATR model utilize an item-based CF to leverage transferred item embeddings. Therefore, it is still questionable whether the item-based CF is the only component to improve performance while transferred item embeddings do not help? To demonstrate the effectiveness of transferred embeddings, we degenerate the NATR via adopting local item embeddings rather than transferred item embeddings in the item-level attention unit. Therefore, it is a kind of *single-domain* method. We name it NATR-local and tune it similarly with NATR to report the best performance.

The compared *cross-domain* baselines are as follows.

- **CMF [32].** *Collective Matrix Factorization* (CMF) decomposes the data matrices of multiple interactions simultaneously while sharing embedding vectors of users or items. Here we factorize two interaction matrices from two domains, sharing embedding vectors of those *bridge items*. We carefully tune the weight of two domains, learning rate and regularizer to report the best performance. It is worth mentioning that a recent study on cross-domain recommendation [43] proposed a method named MPF, which adapted vanilla CMF to a special case where all users and items are all overlapped across domains. Apparently this special setting does not fit our problem where only item can be overlapped and regretfully MPF cannot be adapted to our task.
- **ItemCST [26].** *Coordinate System Transfer* (CST) also assumes that both users and items are overlapped and adds two regularization terms in objective functions. Specifically, the two terms set constraints to the embedding distance in two domains for those overlapped users or items. Thus, CST can be adapted to our problem by only reserving item-side regularization term in our task, and we name it as ItemCST. We tune the learning rate and coefficient of regularization term to report the best performance.

  To conclude, CMF is the state-of-art cross-domain recommendation method while facing with high risk of leaking user privacy since it assumes that all interaction data are fully shared. ItemCST is an adapted method from CST and as it only needs the transferred item embeddings to compute the regularization term of item, it preserves user-relevant data.

We implement the baseline methods and our NATR model in TensorFlow[10] It should be noted that we set the embedding size of all compared methods to be 64, which is a typical setting in literature [14, 17]. Our primary experiments also demonstrate that 64 is an embedding size with enough ability to represent the user and item.

4.2.2 Parameter Settings. To determine the optimal hyper-parameters of the method, we construct a validation set via randomly selecting an interacted item for each test user, which has not been selected as the test item. For our NATR, we initialize parameters with a widely used initialization method proposed in [11]. During the training phase, we intentionally set the negative sampling ratio as 4 to construct mini-batches with size of 256 as described in Section 3.5. To optimize the NATR model, we employ the Adagrad [7] optimizer and search its learning rate within {0.001, 0.002, 0.005, 0.01}. In addition, we tune the λ in Equation 15, which balances the loss and regularization terms, in {1e-2, 1e-3, 1e-4,1e-5,1e-6}. As mentioned before, ItemCST and our NATR only rely on the transferred item embeddings as auxiliary data, and in this paper, without loss of generality, we adopt PMF in the auxiliary domain and carefully tune its learning rate and regularizer to obtain item embeddings.

We first compare the top-K recommendation performance with baseline methods. We investigate the top-K performance with K setting to {1, 2, 5, 10}[11]. As described in the evaluation protocols, we test the performance of a ranking list with 100 items. As such, it is reasonable to choose relatively small K [14]. For every method,

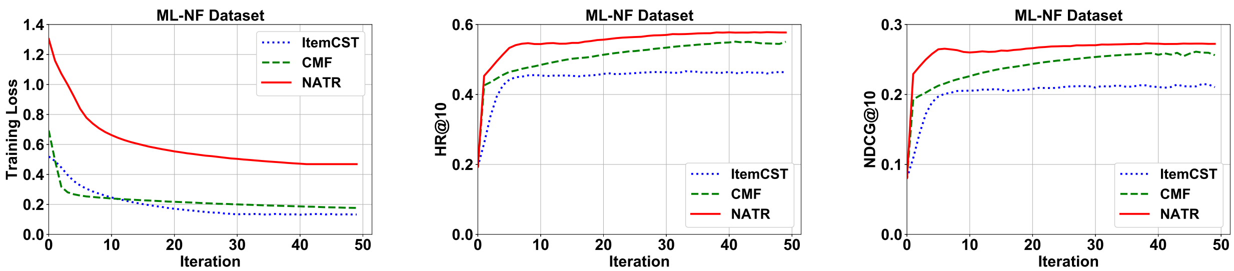
[10] https://www.tensorflow.org

[11] Note that HR@K equals to NDCG@K when setting $K = 1$.

Table 2: Top-K recommendation performance comparison on the ML-NF and TC-IQI datasets (K is set to 1, 2, 5, 10)

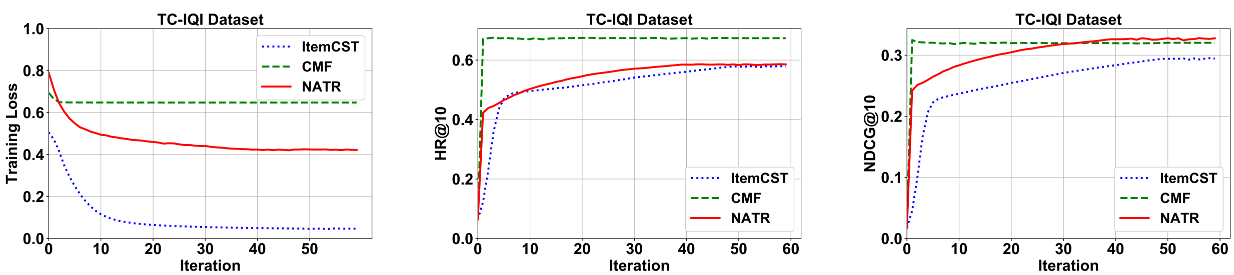
			ML-NF Dataset						
Group	**Method**	**User-relevant Data**	**HR(NDCG)@1**	**HR@2**	**NDCG@2**	**HR@5**	**NDCG@5**	**HR@10**	**NDCG@10**
Cross Domain	**NATR**	**Preserved**	**0.1315**	**0.1976**	**0.1403**	**0.3776**	**0.2110**	**0.5781**	**0.2726**
	ItemCST	Preserved	0.0795	0.1475	0.1005	0.3068	0.1670	0.4846	0.2228
	CMF	Shared	0.1023	0.1903	0.1283	0.3675	0.2025	0.5483	0.2593
Single Domain	NATR-local	Preserved	0.0947	0.1769	0.1253	0.3402	0.1894	0.5183	0.2440
	PMF	Preserved	0.0668	0.1162	0.0796	0.2721	0.1375	0.4494	0.1956
	GMF	Preserved	0.0706	0.1174	0.0816	0.2681	0.1410	0.4284	0.1918

			TC-IQI Dataset						
Group	**Method**	**User-relevant Data**	**HR(NDCG)@1**	**HR@2**	**NDCG@2**	**HR@5**	**NDCG@5**	**HR@10**	**NDCG@10**
Cross Domain	**NATR**	**Preserved**	**0.2010**	0.2660	**0.2104**	**0.4513**	**0.2881**	0.6035	**0.3365**
	ItemCST	Preserved	0.1161	0.2129	0.1445	0.4194	0.2309	0.6079	0.2904
	CMF	Shared	0.1649	**0.3101**	0.2101	0.4499	0.2668	**0.6595**	0.3326
Single Domain	NATR-local	Preserved	0.1677	0.2552	0.1776	0.4214	0.2412	0.5864	0.2948
	PMF	Preserved	0.0848	0.1238	0.0945	0.2291	0.1326	0.3309	0.1694
	GMF	Preserved	0.1584	0.2445	0.1729	0.4101	0.2425	0.6021	0.3029



(a) Training Loss (b) HR@10 (c) NDCG@10

Figure 3: Training loss and testing performance of NATR, ItemCST and CMF in each iteration on ML-NF Dataset.

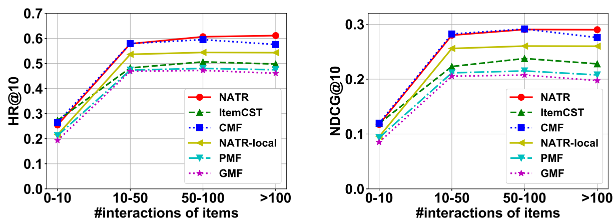


(a) Training Loss (b) HR@10 (c) NDCG@10

Figure 4: Training loss and testing performance of NATR, ItemCST and CMF in each iteration on TC-IQI Dataset.

we carefully tune the hyper-parameters to report the best performance. In Table 2, we report the top-K recommendation performance for the two utilized real-world datasets. We compare our proposed NATR method with three single-domain baselines and two cross-domain ones. From these results, we have the following observations:

- **NATR significantly improves recommendation performance in the target domain.** 1) For those *single domain* methods that are trained with only interaction data from the target domain, the recommendation performance is relatively poor. PMF achieves the worst performance, which can be explained as the limited representation ability of MF model. 2) NATR-local, a degenerative model of the proposed NATR, outperforms PMF and GMF *w.r.t.* all metrics on the ML-NF dataset and most metrics on the TC-IQI dataset, which justifies the effectiveness of explicitly encoding users' historical interactions. 3) Compared with PMF, GMF, and NATR-local, the proposed NATR outperforms the best of them by 28.26% and 39.11% in HR@10 and NDCG@10 for ML-NF dataset and by 0.23% and 11.09% in HR@10 and NDCG@10 for TC-IQI dataset. It demonstrates that leveraging the item embeddings from the auxiliary domain enhances the recommendation quality in the target domain, which



(a) HR@10 (b) NDCG@10

Figure 5: Performance of all methods on items with different number of interaction records on ML-NF dataset.

further indicates that the proposed NATR is a promising solution for the cross-domain recommendation task.

- **NATR performs even better than those cross-domain methods with the risk of leaking user privacy.** We can observe that our proposed NATR model achieves the best performance compared to CMF and ItemCST regarding every evaluation metrics in the NF-ML dataset. For HR@10 and NDCG@10, NATR outperforms the best of them by 18.94% and 4.94%, respectively. We guess the reason of such results is that a joint training with data from two domains (CMF and ItemCST) might distract the loss during the optimization and converge at a status which balances the two domains rather than the optimal status of the target domain. We leave further investigations at the future work. To further study how these methods perform, we present the training loss and testing performance in each interaction in Figure 3 (for ML-NF dataset) and Figure 4 (for TC-IQI dataset). For every method in the two figures, we report the best parameter settings. For both datasets, all methods achieve stable performance after about 50 iterations. With fine hyper-parameter tuning to solve over-fitting, our proposed NATR can effectively outperform ItemCST and achieve similar or even better performance than CMF.
- **NATR effectively distills the CF signal encoded in transferred item embeddings.** NATR-local, a degenerative model of our proposed NATR, only utilizes interaction data from the target domain without exploiting transferred item embddings. Specifically, it replace the transferred item embedding in NATR to local item embeddings. On one hand, the experimental results in Table 2 show that NATR-local achieves better performance than GMF, a competitive method for single domain, demonstrating that taking the explicit preferences of users into consideration can improve recommendation performance. On the other hand, NATR outperforms NATR-local on two datasets which means the combining transferred embeddings are better than only a local CF solution on the target domain. This confirms the utility of transferred item embeddings in encoding CF signal from the auxiliary domain.

To summarize, these comparisons on two real-world datasets verify that our proposed NATR model can effectively leverage transferred item embeddings to improve the recommendation performance in the target domain.

4.3 Data Sparsity Problem (RQ2)

As mentioned in the introduction, one of the primary purposes for cross-domain recommendation is to alleviate item data sparsity problem (*i.e.*, items' records are too few) in the target domain. In particular, for those items with few interactions, of which the embeddings cannot be learned well in the target domain itself, transferred embeddings from the auxiliary domain play a bigger role. To study extensively how our proposed NATR model effectively helps to alleviate the item data sparsity issue, we compare the recommendation performance for items with different level of sparsity.

Specifically, we divide the items to several groups according to number of interaction records in the training set. Note that each group have similar number of items, which make the experimental results more reasonable. Then we apply the evaluation protocol, *leave-one-out*, which is the same with above experiments. For each item, its performance is defined as the average of HR@10 and NDCG@10 when it is in the test set. We compare the proposed NATR model with all five baseline methods in Figure 5. From the results, we can observe that when the interaction records of an item become sparser, the recommendation performance will go worse. For example, in the first group, of which each item has been interacted by only 1-10 users, the best performance of those single-domain methods is about only 0.220 for HR@10 and 0.098 for NDCG@10. Fortunately, with the help of the auxiliary domain, cross-domain methods can achieve better performance for those sparse items. Out of these methods, our proposed NATR model can achieve similar performance compared with CMF and better than ItemCST, which verifies that NATR can serve as a competitive cross-domain recommendation method without sharing user-relevant data.

In summary, our NATR model can improve recommendation performance effectively, no matter the historical records of items are sparse or dense. For items with sparser records, the improvement is more evident and meaningful.

4.4 Impact of Attention Mechanism (RQ3)

In NATR, we utilize the attention mechanism to solve two primary challenges. First, a domain-level attention unit is applied to distill useful signals from transferred item embeddings adn integrate them into the target domain. Second, an item-level attention unit is adopted to handle varying importance of transferred item embeddings. An intuitive question is whether the designed attention unit can really help in our model?

To answer it, we conduct experiments on two degenerative methods of NATR, in which two utilized attention network components are replaced by the simple operation of pooling (*i.e.*, average summation), respectively. We adopt the same evaluation methods with above experiments, and the performance comparison is shown in Table 3. We can observe that when removing item-level attention unit, the top-10 recommendation performance become poor, which

Table 3: Impact of Attention Mechanism.

Dataset	ML-NF	
Methods	**HR@10**	**NDCG@10**
NATR	0.5781	0.2726
Without Item-level Attention	0.5624	0.2655
Without Domain-level Attention	0.5669	0.2722

demonstrate the challenge of varying importance and our attention solution can address it. Besides, removing domain-level attention unit also make the recommendation performance worse. That is, the auxiliary domain cannot be considered as the same with the target domain simply. To conclude, the experimental results demonstrate the necessity of our two specially designed attention units.

In conclusion, extensive experiments on two real-world datasets verify the efficacy of our proposed model, and further studies demonstrate that our model can alleviate data sparsity problem. Moreover, the utility of our specially designed attention networks components is verified.

5 RELATED WORK

In this paper, we propose a solution for cross-domain recommendation without sharing user-relevant data based on neural networks. The closed related work can be divided into cross-domain recommendation and neural network based recommendation.

Cross-domain Recommendation To alleviate cold start and data sparsity issue, cross-domain recommendation is a typical solution which takes data from multiple domains into consideration. With the help of the auxiliary domain, cross-domain recommendation methods can achieve better performance (*i.e.* recommendation accuracy) than single-domain ones. Approaches of cross-domain recommendation can be broadly classified into two types: collaborative and content-based.

Collaborative cross-domain recommendation refers to those approaches utilizing interaction data (rating history, for example) from two domains. Ajit *et al.* [32] proposed a MF based model, CMF (Collective Matrix Factorization), which assumes a common global user factor matrix for all domains, and it factorizes matrices from multiple domains simultaneously. Li *et al.* [22] proposed a model named CBT (Code Book Transfer) which builds a matrix named codebook to represent cluster-level rating pattern and this codebook is shared by two domains. A recent study [43] considered a special task in which both users and items are overlapped, and they proposed a MF based model which assumes part of the user embeddings and whole item embeddings are shared across domains. With a similar setting, Man *et al.* [23] proposed a neural method which employs multi-layer perceptron to adapt user and item embeddings between two domains. Pan *et al.* [26] utilize auxiliary interaction data with a regularization term concerned with overlapped user and item in objective function in MF model. Another category of cross-domain recommendation models is content-based ones, which sharing attributes of user or items from auxiliary domain [1, 8, 45]. Agarwal *et al.* [1] proposed a MF based model in cross-domain recommendation when multi-modal user profiles are available. Elkahky *et al.* [8] transformed user profile and item attributes to dense vectors through deep neural network and matched them in latent space. Zhang *et al.* [45] utilize textual, structure and visual knowledge of items as auxiliary domain to aid building item embedding.

In this paper, we focus on collaborative cross-domain recommendation with auxiliary interaction data, a widely used setting in literature. Specifically, our problem is a typical system-level cross-domain recommendation task, where same items are shared across domains, according to the definition in two surveys [2, 21]. There is a common assumption in existing methods that the whole interaction data can be fully shared across domains, which has the risk of leaking user privacy since various domains may be operated by different companies. In this work, we advocate a more realistic settings that only item-side data can be shared.

Neural Network Based Recommendation. Salakhutdinov *et al.* [30] proposed RBM (Restricted Boltzmann Machines) to predict explicit ratings, which was the first work to apply neural networks to recommender systems. Recently, similar as the research field of CV and NLP, neural networks have achieved great success in recommender systems. Some works [14, 36, 42, 44, 46] relied on neural networks to learn to match function between users and items directly, which can be regarded as the extension of traditional collaborative filtering approaches. He *et al.* [14] proposed a general neural architecture for collaborative filtering, which learns the user-item interaction function via generalized matrix factorization and multi-layer perceptrons. Zhang *et al.* [46] mapped user and item to Hamming space and obtained matching score via neural networks. Tay *et al.* [36] proposed a relational-translation based match function to learn from interactions. Yang *et al.* [42] utilized neural networks to match user-item interaction and user-user relation simultaneously to perform a social recommendation task. Ying *et al.* [44] relied on graph conventional network for collaborative filtering to match huge mount of users and items in real recommender systems. Gao *et al.* [9] introduce multi-task learning to nueral networks to solve the task of recommendation with users' multiple types of behaviors. Besides, some works utilized neural networks to extract the auxiliary information and features in recommender systems, such as textual [25, 47], visual [3, 38], video [5], and hybrid [48].

In this paper, we propose a novel neural model to solve the challenges of extracting useful knowledge from item embedding of the auxiliary domain with the power of neural networks in representative learning.

6 CONCLUSION

In this work, we present a new cross-domain recommendation solution, which can avoid user privacy leakage by transferring only item embeddings from the auxiliary domain. To better exploit the transferred item embeddings, we propose a neural network method named NATR, combining item-level and domain-level attention mechanisms to address the challenges in cross-domain learning. We conduct extensive experiments on two real-world datasets, demonstrating that our NATR method can improve the recommendation performance of the target domain by 18.94%. To the best of our knowledge, this is the first work that concerns user privacy in cross-domain recommendation, and presents a sound solution to exploit the predictive signal without sharing any user-relevant information.

There are three points about this work that we plan to address in future. First, although our method has taken the dimensionality discrepancy into consideration via a dimension-adaption fully-connected layer, we only empirically verify its effectiveness when two domains have the same embedding size. As such, we will study how different sizes of transferred embeddings would affect the recommendation performance. Second, we will study scalability of our method in industrial scenario where amount of users and items are very huge. Last, since this work only focuses on collaborative cross-domain recommendation, we will study the similar task in content-based scenario.

ACKNOWLEDGMENTS

This work was supported in part by The National Key Research and Development Program of China under grant 2017YFE0112300, the National Nature Science Foundation of China under 61861136003, 61621091 and 61673237, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology. This research is also part of NExT++ research which is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.

REFERENCES

[1] Deepak Agarwal, Bee-Chung Chen, and Bo Long. 2011. Localized factor models for multi-context recommendation. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 609–617.

[2] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. Cross-domain recommender systems. In *Recommender systems handbook*. Springer, 919–959.

[3] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 335–344.

[4] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Deep Learning for Recommender Systems Workshop (DLRS)*. 7–10.

[5] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *ACM Conference on Recommender Systems (RecSys)*. 191–198.

[6] Jingtao Ding, Guanghui Yu, Xiangnan He, Yuhan Quan, Yong Li, Tat-Seng Chua, Depeng Jin, and Jiajie Yu. 2018. Improving Implicit Recommender Systems with View Data.. In *International Joint Conferences on Artificial Intelligence (IJCAI)*. 3343–3349.

[7] John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research (JMLR)* 12, Jul (2011), 2121–2159.

[8] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *International World Wide Web Conference (WWW)*. 278–288.

[9] Chen Gao, Xiangnan He, Dahua Gan, Xiangning Chen, Fuli Feng, Yong Li, Tat-Seng Chua, and Depeng Jin. 2019. Neural Multi-Task Recommendation from Multi-Behavior Data. In *International Conference on Data Engineering (ICDE)*.

[10] Chen Gao, Chao Huang, Yue Yu, Huandong Wang, Yong Li, and Depeng Jin. 2019. Privacy-preserving Cross-domain Location Recommendation. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *International Conference on Computer Vision (ICCV)*. 1026–1034.

[12] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. 2018. Adversarial personalized ranking for recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 355–364.

[13] Xiangnan He, Zhenkui He, Jingkuan Song, Zhenguang Liu, Yu-Gang Jiang, and Tat-Seng Chua. 2018. NAIS: neural attentive item similarity model for recommendation. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* (2018).

[14] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *International World Wide Web Conference (WWW)*. 173–182.

[15] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. 2013. Personalized recommendation via cross-domain triadic factorization. In *International World Wide Web Conference (WWW)*. 595–606.

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

[17] Santosh Kabbur, Xia Ning, and George Karypis. 2013. Fism: factored item similarity models for top-n recommender systems. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 659–667.

[18] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 426–434.

[19] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009).

[20] Wenqiang Lei, Xisen Jin, Min-Yen Kan, Zhaochun Ren, Xiangnan He, and Dawei Yin. 2018. Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. 1. 1437–1447.

[21] Bin Li. 2011. Cross-domain collaborative filtering: A brief survey. In *International Conference on Tools with Artificial Intelligence*. 1085–1086.

[22] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction.. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, Vol. 9. 2052–2057.

[23] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: an embedding and mapping approach. In *International Joint Conferences on Artificial Intelligence (IJCAI)*. 2464–2470.

[24] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Annual Conference on Neural Information Processing Systems (NIPS)*. 1257–1264.

[25] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based news recommendation for millions of users. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1933–1942.

[26] Weike Pan, Evan Wei Xiang, Nathan Nan Liu, and Qiang Yang. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *The AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 10. 230–235.

[27] Chanyoung Park, Donghyun Kim, Jinoh Oh, and Hwanjo Yu. 2017. Do also-viewed products help user rating prediction?. In *International World Wide Web Conference (WWW)*. 1113–1122.

[28] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *The Conference on Uncertainty in Artificial Intelligence (UAI)*. 452–461.

[29] Christian Robert. 2014. Machine Learning, a probabilistic perspective. *CHANCE* 27, 2 (2014), 62–63.

[30] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. In *International Conference on Machine Learning (ICML)*. 791–798.

[31] Andriy Shepitsen, Jonathan Gemmell, Bamshad Mobasher, and Robin Burke. 2008. Personalized recommendation in social tagging systems using hierarchical clustering. In *ACM Conference on Recommender Systems (RecSys)*. 259–266.

[32] Ajit P Singh and Geoffrey J Gordon. 2008. Relational learning via collective matrix factorization. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 650–658.

[33] Peijie Sun, Le Wu, and Meng Wang. 2018. Attentive recurrent social recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 185–194.

[34] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning (ICML)*. 1139–1147.

[35] Jie Tang, Sen Wu, Jimeng Sun, and Hang Su. 2012. Cross-domain collaboration recommendation. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1285–1293.

[36] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Latent relational metric learning via memory-based attention for collaborative ranking. In *International World Wide Web Conference (WWW)*.

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems (NIPS)*. 5998–6008.

[38] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *International World Wide Web Conference (WWW)*. 391–400.

[39] Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. 2018. Tem: Tree-enhanced embedding model for explainable recommendation. In *International World Wide Web Conference (WWW)*. 1543–1552.

[40] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Attentional factorization machines: learning the weight of feature interactions via attention networks. In *International Joint Conferences on Artificial Intelligence (IJCAI)*. 3119–3125.

[41] Jun Xu, Xiangnan He, and Hang Li. 2018. Deep learning for matching in search and recommendation. *International World Wide Web Conference (WWW)* (2018).

[42] Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. 2017. Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

[43] Chunfeng Yang, Huan Yan, Donghan Yu, Yong Li, and Dah Ming Chiu. 2017. Multisite user behavior modeling and its application in video recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 175–184.

[44] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

[45] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

[46] Yan Zhang, Hongzhi Yin, Zi Huang, Xingzhong Du, Guowu Yang, and Defu Lian. 2018. Discrete deep learning for fast content-aware recommendation. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 717–726.

[47] Lei Zheng, Vahid Noroozi, and Philip S Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In *International Conference on Web Search and Data Mining (WSDM)*. 425–434.

[48] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1059–1068.