Heterogenous Information Network Embedding based Cross-domain Recommendation System

Jiangyi Yin
School of electronic and information engineering, Beijing Jiaotong University
Beijing, China
17120156@bjtu.edu.cn

Yuchun Guo
School of electronic and information engineering, Beijing Jiaotong University
Beijing, China
ychguo@bjtu.edu.cn

Yishuai Chen
School of electronic and information engineering, Beijing Jiaotong University
Beijing, China
yshchen@bjtu.edu.cn

***Abstract*—The data sparsity or the cold start is a general bottleneck of recommendation service systems. Two types of state-of-art schemes address this issue via enriching knowledge. Cross-domain recommendation schemes transfer knowledge from another dense domain, but they only explore a single kind of shared knowledge. Recommendation schemes built on heterogeneous information networks (HIN) utilize knowledge implied in meta-paths connecting various objects, these methods, however, investigate meta-paths in a single domain. To improve recommendation performance in a highly sparse scenario, we propose a HIN Embedding based Cross-domain Recommendation (HecRec) framework, which exploits cross-domain information by establishing meta-path based HIN embeddings in both the source and the target domain and conducts personalized recommendation by integrating the obtained HIN embeddings with a rating predictor. To make the best use of cross-domain information and avoid the knowledge confliction between knowledge from different meta-paths observed in real-system datasets, we adopt a concept of "overpass bridge" to integrate the HIN embeddings drawn via different meta-paths. Experiments on two public datasets, i.e., MovieLens and LibraryThing, demonstrate the capacity of HecRec on addressing data sparseness. The recommendation performance of HecRec is improved up to 6.9% in MAE and 5.5% in RMSE compared with the state-of-art schemes in the coldest cases.**

***Index Terms*—cross-domain recommendation, heterogeneous information network, network embedding, cold-start problem**

I. Introduction

Cold-start problem and data sparsity are general problems in recommendation systems. With the boosting of online services, recommendation algorithms are widely used to help users find items on their preference. Most recommendation algorithms infer a user's preference by exploring her review or purchase records for specific items. Obviously, the more rating records, the better understanding of a user's preference or an item's feature. However, this method fails for new users or new items, referred to as the cold-start problem of a recommendation algorithm, which is inevitable in reality, especially for a growing system always with new users or items.

Existing methods for solving this problem are not efficient enough, particularly when the system is highly cold. Recently, there are two types of methods widely considered. The first type of methods turns to use the heterogeneous information network (HIN) with more than two kinds of nodes or links [26], [27] to model and explore more information beyond user's rating or purchase records. With the connection (referred as meta paths) among various nodes in the HIN, the features of the users and items represented with embeddings can be learned and used for recommendation [15]–[18]. An example of the HIN is shown in Fig.1 for a movie recommendation system. Based on the connections showing that user John likes the actor Tom Cruise starred movie "Without Limits", the genre "sci-fi" movie "Interstellar", and the director "James Wan", it maybe inferred that John may like sci-fi movie "Avata" or Tom Cruise starred movie "Ask the Dust" with the additional features of users and movies learning from the HIN topology. But such methods only investigate HIN within the target domain so that the recommendation performance is limited by the datas sparsity. Another kind of approach is cross-domain recommendation which transfers knowledge from a source domain with dense records. For example, in movie and book recommender systems, movies and books may share genres, writers, target audience and so on, and users of movies and books may show similar tastes. Therefore, knowledge of user's preference and item features can be transferred between movie and book recommender systems [4]–[6], [8], [28], [29]. However, such methods only consider a single kind of shared knowledge, so that the contribution of the source domain is limited by the "capacity" of the bridge nodes.

To solve the cold-start problem with better performance, it is rational to reconcile these two methods to conduct HIN embedding based recommendation on cross-domain data. In this way, auxiliary domain data can enrich information for learning better HIN embedding, and multiple meta-paths passing same nodes will widen the "bridge " across domains for transferring more knowledge. Take into account this observation, we propose a novel framework HecRec, which builds a cross-domain HIN to comprehensively utilize various relations among various objects within and across domains for a better recommendation in cold start. We first establish meta-path based HIN embeddings within and across the target domain and then apply the integrated HIN embeddings to a rating predictor to conduct personalized recommendation.

However, a challenge comes accordingly when achieving

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00060

this combination since the existing HIN embedding based recommendation algorithm for single domain fails to be extended to the cross-domain situation straightly. Shi et al. [3] learn distinct HIN embeddings according to different meta-paths and map the embeddings into one vector according to mapping function in their single domain recommender. However, our experiments show that, introducing information from auxiliary domain by directly adding cross-domain meta-paths and processing the HIN embeddings in the similar way cannot improve but even degrade the performance when the data get denser. This result indicates that, there exists knowledge confliction between different kinds of meta-paths drawn across domains. The confliction arises because some types of nodes sit on more than one kind of meta-paths and the feature vector learned from distinct meta-path represents knowledge on particular relation. Fusing such conflicting feature vectors directly may compromise and offset the knowledge.

To make the best use of cross-domain information and avoid the knowledge confliction, we adopt a concept of "overpass bridge" to integrate the HIN embeddings, which can maintain the knowledge of each single meta-path without interfering with each other in the process of transferring it across domains.

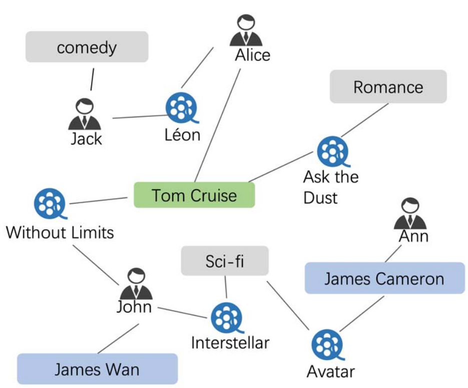

Fig. 1. A HIN constructed according to a movie recommendation system which contains five kinds of objects (user, movie, director, actor, genre) and links between every two kinds of them.

The contribution of our work can be summarized into three parts:

- We propose a novel cross-domain recommendation framework, HecRec, based on integrating cross-domain meta-path HIN embeddings to improve personalized recommend performance, especially for the cold-start situation.
- We find out the knowledge confliction from different meta-paths across domains, and adopt a concept of "overpass bridge" to integrate HIN embeddings effectively for better recommendation performance.
- Experimental results on two public datasets, i.e., MovieLens and LibraryThing, show that our algorithm can efficiently addressing data sparseness problems. For instance, it obtains performance gaining 2.7% improvement in MAE and 2.8% in RMSE compared with the state-of-art schemes, and improves up to 6.9% in MAE and 5.5% in RMSE in the most cold case.

The remainder of the paper is organized as follows. Section II introduces related work including HIN embedding and cross-domain recommendation. Section III presents our proposed method. Section IV shows a series of experiments. Section V draws conclusions and future directions.

II. Related Work

~~In this part, we will review the related researches from two different aspects, HIN embedding based recommendation and cross-domain recommendation.~~

~~Consider the various characteristics and rich semantics in the system, researchers tend to use HIN to present the system and recommend items according to the node embeddings. However, different from the common bipartite networks, which only contains users and items, it is difficult to analyze the complex relations among various kinds of nodes. Chang et al. [21] design a deep embedding algorithm for networked data, which uses a nonlinear multilayered embedding function to capture the complex interactions between the heterogeneous data in a network. Dong [1] formalizes meta-path based random walks to construct the heterogeneous neighborhood of a node and leverages a heterogeneous skip-gram model to perform node embeddings. Yu et al. [12] combine rating data and social information to construct HIN and learn recommend embeddings after finding implicit friends using meta-path based embedding method. Shi et al. [3] propose a recommend approach, HERec, which get embeddings according to each one of meta-paths and then integrate them into MF framework after mapping them into one embedding according to a mapping function. These methods have shown obvious advantage of HIN for recommendation system in capture auxiliary data. However, to our knowledge, there is no existing approach applying it to cross-domain recommendation system and improve performance by introducing more available information from other domains.~~

~~As another direction of solution to mitigate the problem of data sparsity, cross-domain recommendation primarily focuses on how to link source and target domain, and transfer knowledge via the linkage. There exist several ways to connect different domains proposed in earlier researches. Chen et al. [8] propose a transfer learning algorithm TLRec, which exploits the overlapped users and items as a bridge to link different domains and implements knowledge transfer. Li et al. [12] proposed a method of CBT. This model as well as the extended versions [10], [11] generated the user rating pattern, called as codebook, from the auxiliary data and then transferred it to the target domain. However, like many traditional recommendation methods, CBT still explore the information of the rating data. Some researchers notice the huge value hidden in other information in the system like tags given by users. TagCDCF, proposed in [5], is based on an intuitive idea that users with same tastes tend to attach similar tags to items. It uses common tags as bridges to apply the auxiliary information~~

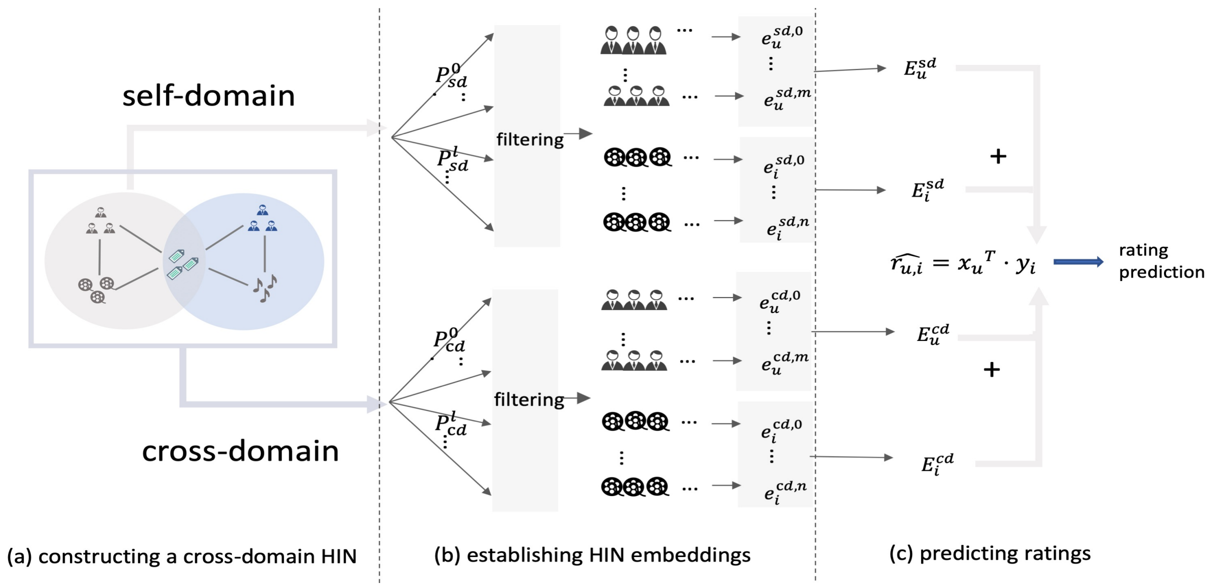

Fig. 2. Proposed framework

~~to analyze user's performance and is able to work well without common users or items. Hao et al. [6] proposed an extension version of TagCDCF, which integrate the rich information contained in domain dependent tags, not only the ones shown in both domains. However, all of the cross-domain methods above explore only one particular kind of shared knowledge, which is not enough. In our proposed model, we establish node embedding according to different meta paths cross domains so that the information can be utilized comprehensively.~~

III. Methodology

In this section, we will present the cross-domain HIN recommendation scheme, HecRec. To show how HecRec works, we take a book recommendation system as a target domain and a movie system as an auxiliary domain. In the beginning of this section, we overview the framework of HecRec. Then we will describe its three components in detail.

A. HecRec framework

To make use of the knowledge implied in multiple meta-paths across domains, we propose the HIN Embedding based Cross-domain Recommendation (HecRec) framework as shown in Fig.2. It consists of 3 components, a) constructing a cross-domain HIN, b) establishing HIN embeddings, and c) predicting ratings based on HIN embeddings. Firstly, to utilize the information from the auxiliary domain, we need to build a cross-domain HIN on the basis of single domain HINs with some common types of nodes as connection. Then we establish HIN embeddings over meta-paths within and beyond the target domain, and further integrate such original HIN embeddings. Finally, we extend the classic matrix factorization [24], [25] recommendation algorithms to incorporate the integrated HIN embeddings to predict user's rating. We will elaborate each components in the following subsections respectively.

B. Building the cross-domain HIN

To build the cross-domain HIN, we need to connect the HIN of the target domain and that of the auxiliary one. HIN in a single domain can be established following the *Definition 1*. The critical problem is to discover linkage between domains to connect the HINs of each single domain.

Based on the data analysis, we choose user-contributed tags as the bridge. It is reasonable to assume that different domains share a quite ratio of user-contributed tags as long as the domains are indeed related. As in our movie and book recommender systems, movies and books may share genres, writers, target audience and so on. These shared objects will appear in the user-contributed tags. Besides, tags will contains many other information that can express users' feeling about the items, and the feelings are exactly what is important when it comes to personalized recommendation.

Based on these assumptions, we can firstly represent each of the domains into a single-domain HIN which contains three types of node (users, items, tags) and then the connect the two networks into a cross-domain HIN like shown in Fig.3. The obtained cross-domain contains five types of nodes (U_t, I_t, T, U_a, I_a) and six types of links.

~~To facilitate further analysis, we will formulate the definition of HIN in a single domain and that across domain respectively as follows.~~

~~*Definition 1:* Heterogeneous Information Network [22]. A HIN is denoted as~~ $\cancel{\mathcal{D} = \{\mathcal{A}, \mathcal{R}\}}$ ~~in which~~ $\cancel{\mathcal{A}}$ ~~denotes the object set and~~ $\cancel{\mathcal{R}}$ ~~denotes the link set in the network, each node~~ $\cancel{a}$ ~~and each link~~ $\cancel{r}$ ~~is associated with a mapping function~~ $\cancel{\phi(a) \to T_a}$ ~~and~~ $\cancel{\phi(r) \to T_r}$~~, where~~ $\cancel{T_a}$ ~~and~~ $\cancel{T_r}$ ~~denotes the type of~~ $\cancel{a}$ ~~and~~ $\cancel{r}$~~.~~ $\cancel{|T_a| + |T_r| > 2}$

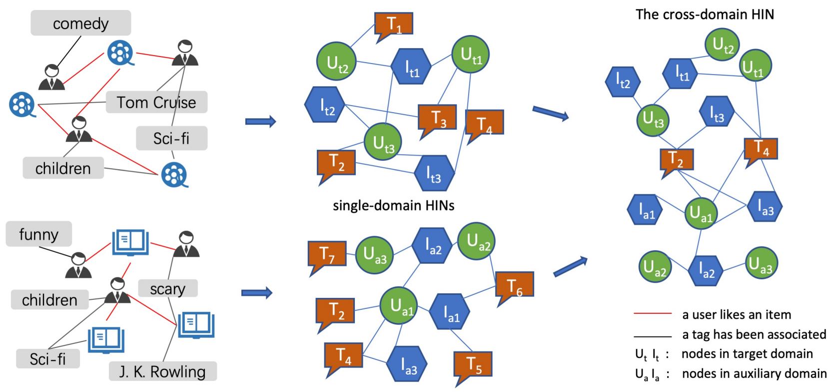

Fig. 3. The procedure of constructing the cross-domain HIN bridged by common user-contributed tags

C. Learning Embeddings across domains

Learning node representation from the cross-domain HIN is the most important pillar in our approach. The learned embeddings are expected able to summarize information from both the target and auxiliary domains.

Different with the single-domain recommendation based on HIN embedding method, we design meta-paths not only in the target domain but also cross domains to fully capture the information in the cross-domain HIN. Moreover, according to our observation, we find out the confliction among original HIN embeddings from distinct meta-paths, and adopt a concept of "overpass bridge" to integrate them.

This part mainly contains three steps. Firstly, we design meta-paths within and beyond the target domain respectively, and conduct random walks based on schemas of the meta-paths. Then we learn original HIN embeddings, using obtained node sequences accordingly. Finally, we process the original HIN embeddings by a personalized and non-linear fusion function and integrate them in the "overpass" way.

1) Random-Walks along Meta-paths: To capture complex and diverse connections between objects, we define a set of meta-paths both within and beyond the target domain and conduct random walks along them to get node sequences. We follow the definition of meta-path given in [19].

Definition 2: Meta-path. A meta-path ρ is defined as a schema in the form of $\mathcal{A}_1 \overset{\mathcal{R}_1}{\rightarrow} \mathcal{A}_2 \overset{\mathcal{R}_2}{\rightarrow} \cdots \overset{\mathcal{R}_l}{\rightarrow} \mathcal{A}_{l+1}$, which describes a composite relation $\mathcal{R}_1 \circ \mathcal{R}_2 \circ \cdots \circ \mathcal{R}_l$ between start node type $\mathcal{A}_1$ and end node type $\mathcal{A}_{l+1}$, where $\circ$ denotes the composition operator on relation.

We design eight different types of meta-paths as shown in TABLE I, half of which stay within the target domain and the other half extend across domains. As bridge nodes linking two domains, tags convey extra knowledge from the auxiliary domain to alleviate cold-start problem in the target domain. Hence 6 types of the meta paths we choose contain tag nodes.

These designed meta-paths can help us find pairs of entities that are actually similar but distant with each other on the user-item bipartite network. For instant, with the meta-path defined as p^1_{cd}, take U_{t3} in Fig.4 as a start node, after 4 steps, $U_{t3} - T_4 - U_{a1} - T_2 - U_{t4}$, the walk reached U_{t4}, which is inferred have relation with U_{t3} as described in TABLE I. The choice of meta-path is flexible to suit the specific dataset and situation.

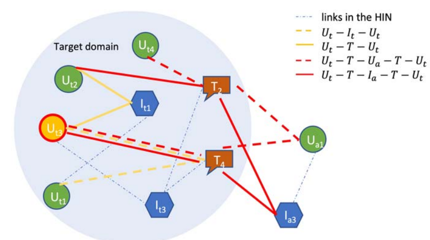

Fig. 4. An illustractive examples of meta-paths. Take U_{t3} as a start node to generate node sequences based on the four meta-paths for users, then we get two self-domain node sequences colored by yellow and two cross-domain node sequences colored by red. The node T_4 acts as a bridge node to reach the auxiliary domain and used in three of the sequences.

Then we conduct random walks starting from each user or item node in the target domain based on all related meta-paths and get amounts of node sequences respectively. Given a meta-path, the node sequences generated by selecting one of the neighbors of the last node in every step. The transition probability of step t is defined according to the schema ρ of the meta-path $\mathcal{A}_1 \overset{\mathcal{R}_1}{\rightarrow} \mathcal{A}_2 \overset{\mathcal{R}_2}{\rightarrow} \cdots \overset{\mathcal{R}_n}{\rightarrow} \mathcal{A}_{n+1}$:

$$P(n_{t+1} = x | n_t = a, \rho) = \begin{cases} \left| \frac{1}{\mathcal{N}^{A_{t+1}}(a)} \right|, (a, x) \in \mathcal{R} \text{ and } \phi(x) = A_{t+1}; \\ 0, \qquad \text{otherwise}, \end{cases} \tag{1}$$

where n_t is the n^{th} node of the sequence and $\mathcal{N}^{A_{t+1}}(a)$ is the neighbors of node a, which is of type $\mathcal{A}_{t+1}$.

Since the obtained sequences are used to learn node embeddings and we only need the embeddings of users and items in the target domain. So after getting the sequences, we filter the nodes with different type from the start one. For example, if we get an original sequence $U_{t1} - I_{t1} - U_{t3} - I_{t2} \cdots$ based on p^0_{sd}, we will remove the item nodes and get a sequence as $U_{t1} - U_{t3} \cdots$.

TABLE I
META-PATHS DESIGNED FOR LEARNING HIN EMBEDDINGS.

	Path	Schema	Describtion
Self-domain	p^0_{sd}	$U_t - I_t - U_t$	Users who have liked the same item
	p^1_{sd}	$U_t - T - U_t$	Users who have used the same tag
	p^2_{sd}	$I_t - U_t - I_t$	Items which have been liked by the same user
	p^3_{sd}	$I_t - T - I_t$	Items which have been tagged the same word
Cross-domain	p^0_{cd}	$U_t - T - I_a - T - U_t$	Users who have used words which are tagged to the same item in the auxiliary domain
	p^1_{cd}	$U_t - T - U_a - T - U_t$	Users who have used words which are used by the same user in the auxiliary domain
	p^2_{cd}	$I_t - T - I_a - T - I_t$	Items which have been tagged by words which are tagged to the same item in the auxiliary domain
	p^3_{cd}	$I_t - T - U_a - T - I_t$	Items which have been tagged by words which are used by the same user in the auxiliary domain

2) Embeddings of Distinct Meta-paths: After getting the node sequences based on distinct meta-paths, we learn the original HIN embeddings by skip-gram [1]. Specifically, we get the context $\mathcal{N}_u$ of node u by selecting the co-occurrence nodes in a pre-defined window size, and then learn the representation by optimizing the objective as:

$$\max_f \sum_{u \in \mathcal{A}} \log \mathcal{P}r(\mathcal{N}_u | f(u)) \tag{2}$$

in which the function $f : u \rightarrow f(u)$ mapping node u to a low dimension embedding. For details of this algorithm please refer to [1].

Different with single domain recommendation algorithm [3], the original HIN embeddings in our work also capture information from the auxiliary domain. Specificly, each user or item in target domain has four embeddings, two obtained from the target domain and the others get cross domains.

3) Integration of Original HIN Embeddings: To facilitate the recommendation algorithm, it is desired to have an integrate of the original HIN embeddings for each user and item in the target domain. We find that the confliction of knowledge among meta-paths prevents the integration scheme of HIN embedding in the single domain recommendation framework [3] to be effective in the cross domain scenario. Therefore we propose an overpass-like integration of original HIN embeddings to accommodate knowledge from different domains.

Shi et al. [3] adopt a personalized non-linear function $g(\cdot)$ to fuse the learned node embeddings. Specifically, given a node $a \in \mathcal{A}$, we obtain a set of original HIN embeddings $\{e_a^n\}^{|\mathcal{P}|}$ where $\mathcal{P}$ denotes the set of meta-paths for users or items and e_a^n denotes the embedding w.r.t the n_{th} meta-path. The fuction $g(\cdot)$ defined as followed:

$$g(\{e_a^n\}) = \sigma(\sum_{n=1}^{|\mathcal{P}|} \omega_a^n \sigma(M^n e_a^n + b^n)) \tag{3}$$

where $M \in \mathbb{R}^{D*d}$ and $b \in \mathbb{R}^d$ are the transformation matrix and bias vector w.r.t the n_{th} meta-path, w_a^n is the node's preference weight on meta-paths, σ is the sigmoid function used to introduce a two-layer nonlinear relationship, which is flexible to extend.

However, this way of fusing original HIN embeddings will degrade when a type of node is reused in different kinds of meta-paths. As shown in Fig.4, the node T_4 appears on three of the 4 sequences, which is because the tags work as the bridge so that it stays in most of the cross-domain meta-paths. As a consequence, node sequences along cross-domain meta-paths and even some sequences from self-domain meta-paths may overlapping on some segments. Obviously, these shared segments convey different knowledge over different meta-paths. Straight fusing the original HIN embeddings together may compromise and offset the distinct knowledge implied by distinct meta paths, which fails to contribute but even damage the performance as shown in experiments over real world datasets. We notice that, the confliction happens not only in the cross-domain meta-paths, but also in the self-domain ones, which just not that serious because the bridge nodes are not badly needed in the self-domain meta-paths so that the segments are less shared.

In order to make the best use of cross-domain information and avoid the knowledge confliction, we adopt the conception of "overpass bridge" in the procedure of transforming and integrating the original HIN embeddings. We firstly apply the function $f(\cdot)$, which introduces personalized factor and non-linear functions, to transform the original HIN embeddings, then we concatenation them. Moreover, because that knowledge from different source works in different way, we process self-domain and cross-domain embeddings separately. In this way, all of the meta-paths just work like the overpass bridge, which is able to make sure that roads work smoothly independently. The integration functions are defined as:

$$e_u^{sd,n} = f(\{e_a^n\}_{n=i}^{|\mathcal{P}_{sd}|}) = \sigma(w_u^{n,sd} \sigma(M^{sd,n} \{e_a^n\}_{n=i}^{|\mathcal{P}_{sd}|} + b^{sd,n})) \tag{4}$$

$$E_u^{sd} = [e_u^{sd,0}, e_u^{sd,1} \cdots, e_u^{sd,N-1}] \tag{5}$$

$$e_u^{cd,n} = f(\{e_a^n\}_{n=i}^{|\mathcal{P}_{cd}|}) = \sigma(w_u^{n,cd} \sigma(M^{cd,n} \{e_a^n\}_{n=i}^{|\mathcal{P}_{cd}|} + b^{cd,n})) \tag{6}$$

$$E_u^{cd} = [e_u^{cd,0}, e_u^{cd,1} \cdots, e_u^{cd,M-1}] \tag{7}$$

TABLE II
STATISTICS OF THE DATASETS

	User	Item	Rating	Sparsity	Common Tag
book	4974	4998	493460	0.724%	699
movie	4999	4796	581684	2.426%	

where $e_a^{sd,i}$ denotes the embedding w.r.t the i_{th} self-domain meta-path of user or item a, E_a^{sd} and E_a^{cd} are final HIN embeddings of node a characterize the features from self-domain and cross-domain respectively.

D. Predicting User Ratings with Embeddings

We proposed a rating predictor enhancing traditional matrix factorization model by adding the HIN information. The predicted rating of user u to item i is formulated as:

$$\begin{aligned}\widehat{r_{u,i}} = x_u^{\mathrm{T}} \cdot y_i + ({E_u^{sd}}^{\mathrm{T}} \cdot r_i^{sd} + {r_u^{sd}}^{\mathrm{T}} \cdot E_i^{sd}) \\ + ({E_u^{cd}}^{\mathrm{T}} \cdot r_i^{cd} + {r_u^{cd}}^{\mathrm{T}} \cdot E_i^{cd})\end{aligned} \tag{8}$$

where x_u and y_i are latent factors of user u and item i, E_u^{sd}, E_u^{cd}, E_i^{sd} and E_i^{cd} are the united HIN embeddings introduced in section 4.3, and r_u^{sd}, r_u^{cd}, r_i^{sd} and r_i^{sd} are latent factors to pair with the HIN embeddings respectively, which are also need to be learned like x_u and y_i. The reason of introducing the four latent factors is that we get the original HIN embeddings from separated spaces. So there is no relation between E_u^{sd}, E_u^{cd} and E_i^{sd} , E_i^{cd} like x_u and y_i and we cannot directly multiple them as in MF. Besides, instead of fusing self-domain and cross-main HIN embeddings together, we handle them separately into the rating predictor for the reason of trying to maintain the individual feature getting from the self-domain and cross domains.

We learn the latent factors by minimizing the objective defined as:

$$loss = \sum_{r_{i,j}} (r_{i,j} - \widehat{r_{i,j}})^2 + \lambda(\|x_u\|_2 + \|y_i\|_2 + \|R\|_2 + \|\Phi\|_2) \tag{9}$$

where the first item denotes the loss and second one is regularization terms, in which R is the set of latent factors to pair with HIN embeddings and Φ denotes the factors introduced in the transformation function $f(\cdot)$ including M, b and w. After then, we can predict user's rating and recommend items for them by ranking the personalied ratings.

IV. EXPERIMENT

In this section, we perform experiments on real world datasets to demonstrate the effectiveness of HecRec. We will mainly answer the following questions about our framework: a) Is it useful to introduce cross-domain data into recommendation system? b) Does it work better to follow the conception of "overpass bridge"? c) Can HecRec truly relieve the cold-start problem?

A. Datasets and Performance Metrics

We evaluate our framework on two open datasets commonly used in recommendation researches, the Movelens Dataset (ml-20m)[1] and the LibraryThing Dataset[2]. The movie datasets were created by 138493 users, containing about 20 million ratings of 27278 movies and 465564 tags contributed by users. The book datasets, containing 7279 users and 37,232 items, has 749,401 ratings and 2,056,487 tag assignments. There are more than 2000 common tags in these two domains.

In our experiments, we choose to use the subset of the first 5000 users and 5000 items in both of the domains to come along with our baselines. Besides, the scale of the subsets is almost the largest compared with datasets used in other existing cross-domain recommendation algorithms. The detailed statistics of the used datasets are shown in TABLE II. Obviously, the movie dataset is denser, which is why we choose it as the auxiliary domain to confirm the capacity of HecRec for cold-start problem. The ratings are scaled from 0 to 5, with half star increments. In the stage of getting HIN embeddings, we filter the ratings not larger than 3 in training dataset to make sure the user-item links are postive. And in the stage of learning rating predictor, all the ratings are used to learn the embeddings.

We choose the most widely used metrics, mean absolute error (MAE) and root mean square error (RMSE) to evaluate the performance. The definitions of the metrics are as followed:

$$MAE = \frac{1}{|\mathcal{D}_{test}|} \sum_{i,j \in \mathcal{D}_{test}} |r_{i,j} - \widehat{r_{i,j}}| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{|\mathcal{D}_{test}|} \sum_{i,j \in \mathcal{D}_{test}} (r_{i,j} - \widehat{r_{i,j}})^2} \tag{11}$$

B. Baselines and Settings

To prove the superiority of our approach, we compare the performance of the HecRec with the following approaches. In these selected baselines, UBCF [23], IBCF [13] and PMF [7] are typical traditional recommendation approaches, ETag-iCDCF is a state-of-art approach of cross-domain recommendation, HERec [3] is a novel HIN based recommendation algorithm, and HERec$_{cd}$ is a variant of HERec via straight extension towards cross-domain scenario. It directly add cross-domain meta-paths and process the embeddings as same as HERec. So that we can use this approach to conduct a control experiment to figure out if our "overpass" bridge way to integrate the original embeddings can works better.

In the experiments, we randomly select 80% ratings as training set, and using the remaining 20% ratings as the test set and get average of 10 times results as the final performance. To come along with [6], the length of the random walk is set to 40. And we set the dimension of HIN embeddings and latent factors as 128 and 30 respectively. The regularization parameter λ is set to 0.1. Compared with HERec, the additional

[1]http://www.grouplens.org/node/73
[2]http://www.macle.nl/tud/LT/

TABLE III
PERFORMANCE COMPARISON WITH BASELINES

	IBCF	UBCF	PMF	ETagiCDCF	HERec	HERec$_{cd}$	HecRec
MAE	0.7248	0.6794	0.6797	0.6789	0.6563	0.6562	**0.6384**
RMSE	0.8808	1.0099	0.8788	0.8556	0.8481	0.8482	**0.8244**

cost of computational complexity of our proposed framework is resulted from the concatenation of the HIN embeddings, cause the dimension of the final HIN embedding is N_{path} times of the ones in HERec, where N_{path} is the number of the metapaths. And considering the size of the network, this kind of linear growth of the cost is acceptable.

C. Results and Discussions

The performance of HecRec and the baseline approaches are shown in TABLE III. And we conduct further experiments to compare our approach with HERec and its variant HERec$_{cd}$. To examine the performance of our HecRec algorithm with respect to different extent of data sparsity, we partition various groups of users according to the numbers of their ratings. The distribution of user's rating number is presented in Fig.5. And the results of performance are shown in TABLE IV.

The results shown in TABLE III demonstrate that HecRec significantly outperforms all of the other baselines. It obtains performance gaining 2.7% improvement in MAE and 2.8% in RMSE compared with the state-of-art schemes. Experiments over different user groups with sparse activities verify that the sparser the situation, the powerful the HecRec. It improves up to 6.9% in MAE and 5.5% in RMSE in the first group of users.

Among the baselines, both ETagiCDCF and HERec perform better than traditional schemes showing the value of cross-domain information as well as HIN embeddings. HERec is much more better than ETagiCDCF, which verifies the strength of HIN embeddings of enriching knowledge from connections of multiple objects. However, HERec$_{cd}$, the cross-domain variant version of HERec which strainghtly applys the native HERec in the cross-domain HIN shows almost neglectable improvement in MAE and even become worse in RMSE compared to the native HERec working on single domain, which confirms the significance of propor manner of integration of HIN embeddings in cross-domain situation.

Based on the results of the above experiments, we can answer the three questions raised earlier.

Q1. Is it useful to introduce cross-domain data into recommendation system?

From the results in TABLE III, HecRec performs much better than HERec according to both of the metrics. But we cannot be sure that it is only because of the auxiliary-domain information. However, results showed in TABLE IV about HERec$_{cd}$ and HERec is much more clear. On data of users with less than 10 ratings, HERec$_{cd}$ indeed achieves a better performance, and the only difference between these two baselines is whether use cross-domain meta-paths or not. Therefore, auxiliary information from related domains can truly benefit. Besides, this statement can also been proved according to the results about PMF and ETagiCDCF in TABLE III.

Q2. Does it work better in our "overpass" way?

We conduct experiments on HERec$_{cd}$ to answer this question. In TABLE IV, although HERec$_{cd}$ performs better to HERec on the most sparse data, it achieves hardly improvement as the data becomes denser, and there is nearly no difference on the whole dataset as shown in TABLE III. So, we can say that, with the fusion way to process HIN embeddings in HERec, the contribution of the auxiliary domain is quite limited. However, our proposed method achieves an apparent improvement, which means our "overpass" way to process and integrate HIN embeddings are indeed more suitable for cross-domain recommendation and

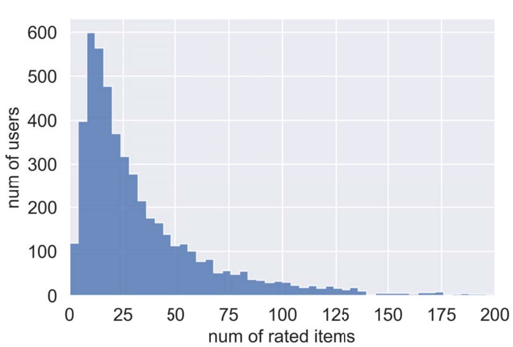

Fig. 5. Distribution of user's rating number

TABLE IV
COMPARISON WITH HEREC ON COLD USER GROUPS

Groups[1]	Metric	HERec	HERec$_{cd}$	HecRec	Improve (%)
(0,10]	mae	0.9017	0.8930	**0.8398**	6.8666
	rmse	1.1640	1.1592	**1.0993**	5.5625
(10,20]	mae	0.6899	0.6870	**0.6803**	1.4050
	rmse	0.9162	0.9127	**0.9077**	0.9310
(20,30]	mae	0.7063	0.7066	**0.7033**	0.4329
	rmse	0.9281	0.9280	**0.9278**	0.0322
ALL	mae	0.6563	0.6562	**0.6384**	2.7301
	rmse	0.8481	0.8482	**0.8244**	2.8025

[1]. The dataset is partitioned into a few parts according to the number of users rating. More ratings, the sub-dataset is more dense.

can works better.

Q3. Can HecRec relieve the cold-start recommendation problem?

In our framework, we solve the problem mainly from two aspects. On one hand, introducing auxiliary domain data can help to extend the range of useful information. On the other hand, applying meta-path based embedding method and adapting "overpass" way to utilize the embeddings can make the bridge wider to transfer more knowledge. The results in TABLE IV shows that, our framework improves about 2.8% compared with the best baseline on the whole dataset, and as data becomes colder, the improvements get more obvious. On data of users with less than 10 ratings, the improvement is up to 6.8% in MAE and 5.5% in RMSE. These results indicate the capacity of HecRec to relieve the cold-start problem.

V. CONCLUSION

In this paper, we propose a novel cross-domain recommendation framework named HecRec, which uses a HIN to comprehensively utilize various cross-domain object and relation information to address the data sparse problem. And we proposed a new conception of "overpass bridge" to integrate HIN embeddings effectively due to the observed knowledge confiliction among different kinds of meta-paths. The experiment showed that, our proposed framework improves up to 2.7% in MAE and 2.8% in RMSE compared with the state-of-art schemes, and achieves improvement up to 6.9% in MAE and 5.5% in RMSE in the most cold case.

As future works, we will involve further investigation in finding worth of large amount of tags not appearing in both domains. And we consider using neural network to replace the inner product in the stage of learning recommendation embeddings and improve performance by learning a more complex relationship between users and items.

REFERENCES

[1] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2017, pp. 135–144.

[2] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in Proc. 24th Int. Conf. World Wide Web, 2015, pp. 1067–1077.

[3] C. Shi, B. Hu, W. X. Zhao, and P. S. Yu, "Heterogeneous information network embedding for recommendation," 2017, arXiv:1711.10730.

[4] B. Li, Q. Yang, and X. Xue, "Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction," Proc. 21st Int'l Joint Conf. Artificial Intelligence, July 2009.

[5] Y. Shi, M. Larson, and A. Hanjalic, "Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering," in UMAP, 2011, pp. 305–316.

[6] P. Hao, G. Zhang, and J. Lu, "Enhancing cross domain recommendation with domain dependent tags," in International Conference on Fuzzy Systems, 2016

[7] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in Proc. Advances Neural Inf. Process. Syst., 2008, pp. 1257–1264.

[8] Z. Lu, W. Pan, E. W. Xiang, Q. Yang, L. Zhao, and E. Zhong, "Selective transfer learning for cross domain recommendation," in Proc. 13th SIAM Int. Conf. Data Mining, 2013, pp. 641–649.

[9] W. Wang, Z. Chen, J. Liu, Q. Qi, and Z. Zhao, "User-based collaborative filtering on cross domain by tag transfer learning," in Proc. 1st International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining, 2012, pp. 10–17.

[10] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," in Proc. 26th Annual International Conference on Machine Learning, 2009, pp. 617–624.

[11] O. Moreno, B. Shapira, L. Rokach, and G. Shani, "Talmud: Transfer learning for multiple domains," in Proc. ACM Conf. Inf. Knowl. Manage, 2012, pp. 425–434.

[12] J.Yu, M.Gao, J. Li, H. Yin, and H. Liu, "Adaptive implicit friends identification over heterogeneous network for social recommendation," in Proc. 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 357-366.

[13] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Itembased Collaborative Filtering Recommendation Algorithms", in Proc. 10th International Conference on the World Wide Web, pp. 285-295, 2001.

[14] J. Huang, X. Cheng, J. Guo, H. Shen, and K. Yang, "Social Recommendation with Interpersonal Influence," in Proc. 19th European Conf. Artificial Intelligence (ECAI '10), 2010. pp. 1113–1133.

[15] H. Yin, H. Chen, X. Sun and H. Wang, "SPTF: A Scalable Probabilistic Tensor Factorization Model for Semantic-Aware Behavior Prediction," in Proc. 17th IEEE International Conference on Data Mining, 2017.

[16] H. Yin, W. Wang, H. Wang, H. Wang, L. Chen, and X. Zhou, "Spatialaware hierarchical collaborative deep learning for poi recommendation," in Proc. IEEE Transactions on Knowledge and Data Engineering, 2017, pp.1-1.

[17] X. Yu et al., "Recommendation in heterogeneous information networks with implicit user feedback," in Proc. 7th ACM Conf. Recommender Syst.,2013, pp. 347–350.

[18] H. Zhao, Q. Yao, J. Li, Y. Song, and D. L. Lee, "Meta-graph based recommendation fusion over heterogeneous information netw- orks," in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2017, pp. 635–644.

[19] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," Proc. 37th Very Large Data Bases, vol. 4, no. 11, pp. 992– 1003, 2011.

[20] X. Yu, X. Ren, Q. Gu, Y. Sun, and J. Han, "Collaborative filtering with entity similarity regularization in heterogeneous information networks," in Proc. 1st Int. Joint Conf. Artif. Intell. Workshop Hetero- geneous Inf. Netw. Anal., 2013.

[21] S. Chang, W. Han, J. Tang, G. J. Qi, C. C. Aggarwal, and T. S. Huang, "Heterogeneous network embedding via deep architectures," in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov- ery Data Mining, 2015, pp. 119–128.

[22] Y. Sun, Y. Yu, and J. Han, "Ranking-based clustering of heterogeneous information networks with star network schema," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 797–806.

[23] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999, pp. 230–237.

[24] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techni- ques for recommender systems," Comput., vol. 42, no. 8, 2009, pp. 30– 37.

[25] Y. Shi, X. Zhao, J. Wang, M. Larson, and A. Hanjalic, "Adaptive diversification of recommendation results via latent factor portfolio," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 175–184.

[26] Y. Sun and J. Han, "Mining heterogeneous information networks: A structural analysis approach," ACM SIGKDD Explorations News- letter, vol. 14, no. 2, pp. 20–28, 2013.

[27] M. Ou, P. Cui, F. Wang, J. Wang, W. Zhu, and S. Yang, "Comparing apples to oranges: A scalable solution with heteroge- neous hashing," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Dis- covery Data Mining, 2013, pp. 230–238.

[28] B. Li, X. Zhu, R. Li, C. Zhang, X. Xue, and X. Wu, "Cross-domain collaborative filtering over time," in Proceedings of the Twenty- Second international joint conference on Artificial Intelligence, 2011, pp. 2293–2298.

[29] B. Li, "Cross-domain collaborative filtering: A brief survey," in Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, 2011, pp. 1085–1086.