Unified Embedding Model over Heterogeneous Information Network for Personalized Recommendation

Zekai Wang[1], **Hongzhi Liu**[1*], **Yingpeng Du**[1], **Zhonghai Wu**[2,3] and **Xing Zhang**[4]
[1]School of Software and Microelectronics, Peking University, Beijing, P.R. China
[2]National Engineering Center of Software Engineering, Peking University, Beijing, P.R. China
[3]Key Lab of High Confidence Software Technologies (MOE), Peking University, Beijing, P.R. China
[4]School of Electronics Engineering and Computer Science, Peking University, Beijing, P.R. China
{tiefblau, liuhz, dyp1993, wuzh, zhx}@pku.edu.cn

Abstract

Most of heterogeneous information network (HIN) based recommendation models are based on the user and item modeling with meta-paths. However, they always model users and items in isolation under each meta-path, which may lead to information extraction misled. In addition, they only consider structural features of HINs when modeling users and items during exploring HINs, which may lead to useful information for recommendation lost irreversibly. To address these problems, we propose a HIN based unified embedding model for recommendation, called HueRec. We assume there exist some common characteristics under different meta-paths for each user or item, and use data from all meta-paths to learn unified users' and items' representations. So the interrelation between meta-paths are utilized to alleviate the problems of data sparsity and noises on one meta-path. Different from existing models which first explore HINs then make recommendations, we combine these two parts into an end-to-end model to avoid useful information lost in initial phases. In addition, we embed all users, items and meta-paths into related latent spaces. Therefore, we can measure users' preferences on meta-paths to improve the performances of personalized recommendation. Extensive experiments show HueRec consistently outperforms state-of-the-art methods.

1 Introduction

Recommender system has became an essential part in various internet services as it can help people to tackle the problem of information overload and find useful information with less time. Recently, heterogeneous information network (HIN) based recommendation models attract much attention due to its advantage of modeling complex information [Yu *et al.*, 2013; Shi *et al.*, 2017]. Most HIN based recommendation models use meta-paths [Sun *et al.*, 2011] to mine the semantic relations between users and items. As different meta-paths can capture different semantics [Jiang *et al.*, 2018; Liu *et al.*, 2019] and result in different recommendation lists [Shi *et al.*, 2015], information extracted from different meta-paths should be fused properly.

To extract and fuse information from different meta-paths, most of existing methods follow a two-step process. First, they explore and extract information from HINs along each meta-path independently. Secondly, information from different meta-paths are fused and then incorporated in a recommendation model. To extract information under a given meta-path, existing methods always use PathSim [Sun *et al.*, 2011] or its variants to measure the semantic relations between users and items [Shi *et al.*, 2015; Han *et al.*, 2018]. However, these methods rely on explicit path reachability, and may fail when path connections are sparse or noisy [Shi *et al.*, 2019]. To fuse information from meta-paths, most methods [Yu *et al.*, 2013; Shi *et al.*, 2016] try to learn a weight vector for meta-paths. However, they always assume the weights on meta-paths are the same for all users without consideration of users' personalized preferences on meta-paths. Recently, [Shi *et al.*, 2019] proposed to learn a separate weight vector on meta-paths for every user. However, this method introduces too many parameters to be learned.

Compared with traditional recommendation model like matrix factorization [Koren *et al.*, 2009], HIN based methods have achieved performance improvement to some extent. However, there still exist two problems for these models. First, they separately model users and items under each meta-path, which assumes users' and items' representations are mutual independent under different meta-paths. This kind of information extraction could be misled and contains some errors, since connections along one meta-path may be sparse or noisy [Shi *et al.*, 2019]. Second, the two-step framework may lead to useful information for recommendation lost irreversibly in the first step. They only aim to model structural features of HINs for information extraction during the first step, then some information which is useful for recommendation but less important for modeling HIN structural features will loss. Moreover, we cannot get the lost information back in the second step, *i.e.* the information loss is irreversible.

To address the first problem, we assume there exist some common characteristics under different meta-paths for each user or item, and these characteristics can be modeled by latent vectors, *i.e.* embeddings. Inspired by the translation based model [Bordes *et al.*, 2013] in the knowledge graphs,

*Corresponding author

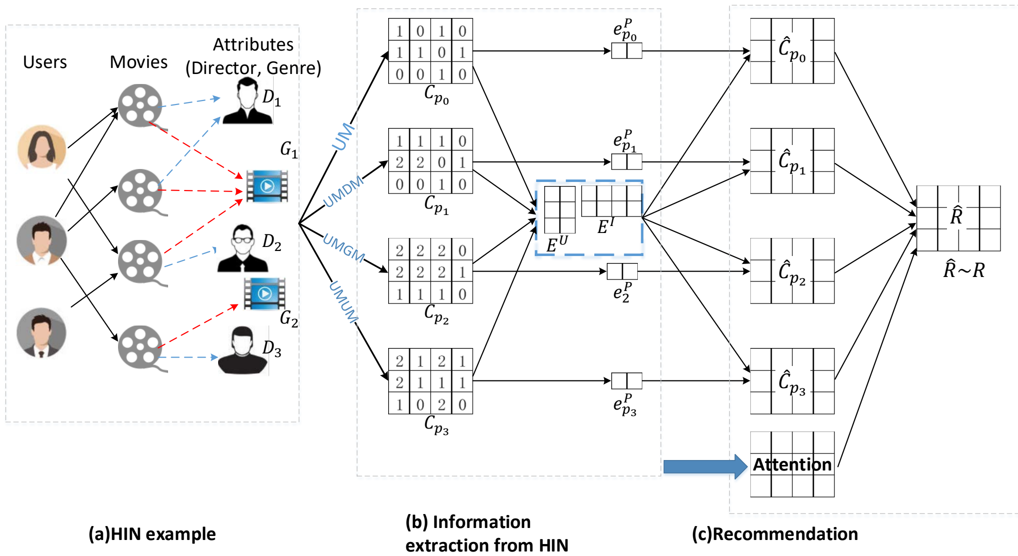

Figure 1: The overall framework of the proposed HueRec approach. UM denotes the meta-path 'User $\rightarrow$ Movie', UMGM denotes the meta-path 'User $\rightarrow$ Movie $\rightarrow$ Genre $\rightarrow$ Movie', and so on. C_{p_i}, and $\widehat{C}_{p_i}$ are the original user-item commuting matrix based on meta-path p_i and the corresponding predicted matrix, respectively. R and $\widehat{R}$ are the original and predicted implicit feedback matrix, respectively. E^U and E^I are unified user and item embedding matrices. $e^P_{p_i}$ is the embedding vector for meta-path p_i and w_{p_i} is the global weight for meta-path p_i.

we embed users and items into one unified latent space, instead of learning a embedding space for users and items independently based on each meta-path. We use data from all meta-paths to learn the unified user and item embeddings, *i.e.* utilizing the interrelation between meta-paths to alleviate the problems of data sparsity and noises on one meta-path. Following this method even if there is little information about a user based on some meta-paths, information from other meta-path is available, which helps to prevent under-fitting or over-fitting.

To address the second problem, we combine the two steps (*i.e.* information extraction and recommendation) into an end-to-end training model to avoid useful information lost in initial phases. Simultaneously optimizing the objectives of both steps can help to distinguish and keep the information useful for recommendation.

In addition, we assume different users have different preferences on meta-paths. Besides users' and items' embeddings, we also learn embeddings of meta-paths themselves, which can reflect characteristics of meta-paths. Based on users' and meta-paths' embeddings, we utilize an attention mechanism [Bahdanau *et al.*, 2015] to measure the preferences of users on meta-paths.

By integrating the above mechanisms, this work presents a HIN based unified embedding model for recommendation, called **HueRec**. The basic idea for the proposed approach is shown in Fig. 1.

The main contributions of this work can be summarized as follow:

1. Based on meta-paths, we propose an end-to-end information exploring and exploiting framework over heterogeneous information network for top-n recommendation.
2. We propose a joint learning method to get unified embeddings of users and items from different meta-paths, so information is fused to some extent in the information extraction step.
3. Based on users' and meta-paths' embeddings, the framework can effectively measure users' preferences on meta-paths using an attention mechanism.
4. We conduct extensive experiments on three real-world datasets, which demonstrate the effectiveness of proposed method.

2 Preliminary

2.1 Heterogeneous Information Network

HIN is a kind of information network involving multiple-typed objects and multiple-typed links, which is defined as follows:

Definition 1 ***Heterogeneous Information Network*** *[Sun* et al.*, 2011]. A HIN is a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an entity type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{A}$ and a link type mapping function $\varphi : \mathcal{E} \rightarrow \mathcal{R}$. $\mathcal{A}$ and $\mathcal{R}$ denote the sets of predefined entity and link types, where $|\mathcal{A}| + |\mathcal{R}| > 2$.*

Due to the complexity of HINs, we use meta-paths to describe the semantic relations between two nodes.

Definition 2 ***Meta-path*** *[Sun* et al.*, 2011]. A meta-path p is defined on a network schema $T_G = (\mathcal{A}, \mathcal{R})$ and is denoted as a path in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$ (abbreviated as $A_1 A_2 \cdots A_{l+1}$), which describes a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between object A_1 and A_{l+1}, where $\circ$ denotes the composition operator on relations.*

In recommender systems, we only concern users' interests on items. So only meta-paths from USER to ITEM are selected in this paper. To measure connections between users to items based on meta-paths, commuting matrix can be used.

Definition 3 ***Commuting Matrix*** *[Sun* et al.*, 2011]. Given a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and its network schema T_G, a commuting matrix C_p for a meta-path $p = (A_1 A_2 \ldots A_l)$ is defined as $C_p = W_{A_1 A_2} W_{A_2 A_3} \ldots W_{A_{l-1} A_l}$, where $W_{A_i A_j}$ is the adjacency matrix between type A_i and type A_j. $C_{i,j,p}$ represents the number of paths instances between object $x_i \in A_1$ and object $y_j \in A_l$ under meta path p.*

2.2 Problem Definition

In this paper, we try to solve the personalized recommendation problem based on HINs. We focus on the common scenario with implicit feedback (e.g. clicks, purchases). Compared with explicit feedback data, implicit feedback data is more widespread and easier to be collected [Liu *et al.*, 2018]. With n users $\mathcal{U} = \{u_1, \cdots, u_n\}$ and m items $\mathcal{I} = \{i_1, \cdots, i_m\}$, we define implicit user feedback matrix $\mathcal{R} \in \mathbb{R}^{n \times m}$ as follows,

$$r_{ij} = \begin{cases} 1, & \text{user } u_i \text{ gives feedback on item } i_j; \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Given a heterogeneous information network $\mathcal{G}$ with user feedback matrix $\mathcal{R}$, we aim to recommend a ranked list of items for each user u_i according to users' preferences on items.

3 The Proposed Model

3.1 Model Overview

In this section, we present the proposed HIN based unified embedding model for recommendation, called **HueRec**. The overall framework of this model is shown in Fig.1. We assume there exist common characteristics of users and items under different meta-paths, and use data from all meta-paths to learn unified users' and items' representations (Section3.2), as shown in Fig. 1(b). We assume users have different preferences on meta-paths and try to learn this kind of preferences to make better personalized recommendation (Section3.3), as shown in Fig. 1(c). In addition, we propose an end-to-end training method for parameter learning to avoid useful information for recommendation lost during separated information extraction phase (Section3.4).

3.2 Embedding of HIN Information

We regard meta-paths between USER and ITEM as contexts for user-item interactions and assume they contain useful information for the predictions of user-item interactions [Hu *et al.*, 2018]. However, the original commuting matrices of these meta-paths may be sparse or noisy. To extract useful information for user and item modeling, we propose to learn the embedding representations of all users, items and meta-paths simultaneously. Therefore, we can use the interrelation information between meta-paths to alleviate the problems of data sparsity and noises on individual meta-paths.

To measure the effectiveness of the embedding representations based on all l meta-paths $\mathcal{P} = \{p_1, \cdots, p_l\}$, we define a loss function for each user-item pair (u, i) as follows,

$$Loss_{u,i}^{HIN} = \frac{1}{l} \sum_{p=1}^{l} Loss_{u,i,p}^{HIN} \cdot w_p, \tag{2}$$

where $Loss_{u,i,p}^{HIN}$ and w_p denote the loss and the corresponding weight of the p-th meta-path, respectively. We assume meta-paths have different importances for recommendation and use weights w_p to control the importances. However, the importance of each meta-path for recommendation is unknown during exploring HINs, we remain it to be defined in Section3.3.

The $Loss_{u,i,p}^{HIN}$ in Eq. 2 represents the loss function of HIN information extraction under the p-th meta-path. We define it based on the sigmoid cross entropy,

$$\begin{aligned} &Loss_{u,i,p}^{HIN} = -ylog(\sigma(\widehat{c}_{u,i,p})) - (1-y)log(1 - \sigma(\widehat{c}_{u,j,p})), \\ &y = \begin{cases} 1, & c_{u,i,p} \geq average(\mathbf{C}_{p,u}); \\ 0, & c_{u,i,p} < average(\mathbf{C}_{p,u}), \end{cases} \end{aligned} \tag{3}$$

where σ is the sigmoid function. $c_{u,i,p}$ denotes the number of path instances between user u and item i under meta-path p, and $\widehat{c}_{u,i,p}$ denotes the corresponding estimated value. $average(\mathbf{C}_{p,u})$ denotes the average of non-zero values in the u-th row of the meta-path commuting matrix C_p.

We assume the path counts $c_{u,i,p}$ are determine by some latent factors of users, items and meta-paths. Therefore, the key for the estimation is to determine the latent representations. Different from previous methods which assume users and items have different latent representations under different meta-paths, we regard meta-paths as contexts and assume there exist unified latent representations for users and items under different meta-paths. Let $\mathbf{e}_u^U \in \mathbb{R}^d$, $\mathbf{e}_i^I \in \mathbb{R}^d$, and $\mathbf{e}_p^P \in \mathbb{R}^d$ denote the latent representations of user u, item i and path p, respectively, where d is the dimension of the latent space. We estimate the path count $c_{u,i,p}$ as follows,

$$\widehat{c}_{u,i,p} = f(\mathbf{e}_u^U, \mathbf{e}_i^I, \mathbf{e}_p^P) = \sum_{q=1}^{d} e_{u,q}^U e_{i,q}^I e_{p,q}^P, \tag{4}$$

where $e_{u,q}^U$ is the q-th bit of the user latent vector $\mathbf{e}_u^U$ and so on.

As a result, the unified embeddings of users and items are shared in all $Loss_{u,i,p}^{HIN}$. For each meta-path p, the HIN information based on p can be extracted into the unified embeddings by minimizing its loss function $Loss_{u,i,p}^{HIN}$. Therefore, by minimizing the aggregated loss function in Equation 2, we can learn common characteristics of users and items on all meta-paths.

3.3 Personalized Preferences on Meta-paths

We assume users' preferences on meta-paths are different. So we need to personalize the weights on meta-paths, and then use them to predict users' interests on items. We utilize an attention mechanism to measure users' personalized preferences on meta-paths. As the data may be sparse for individual users, it may be difficult to directly learn the personalized weights for some users individually. To solve this problem, we propose to integrate data of different users and learns the global weights for meta-paths. So the preference of a user on a meta-path consists of two parts: a global weight and a personalized weight.

We use the weighted sum of scores to predict the interest of a user u on an item i:

$$\widehat{r}_{u,i} = \sum_{p=1}^{l} (w_{u,p} + w_p^{Global})\widehat{c}_{u,i,p}. \tag{5}$$

where $w_{u,p}$ denotes personal preference of a user u on a meta-path p, w_p^{Global} denotes the global weights on meta-paths, and $\widehat{c}_{u,i,p}$ denotes the estimated score *w.r.t.* meta-path p as given in Eq. 4. The global weights w_p^{Global} will be useful for users with little records, whose personalized weight $w_{u,p}$ may be unprecise. Since w_p^{Global} shows global importance of the meta-path p, it can be used as the weight of p in the HIN information extraction, *i.e.* w_p in Eq. 2. We suppose that for any meta-path, its contribution to information extraction and recommendation should be non-negative and the sum of all weights should be equal to 1. To this end, we define the global weight vector on meta-paths $\mathbf{w}^{Global} = (w_1^{Global}; \cdots ; w_l^{Global})$ as follow:

$$\mathbf{w}^{Global} = softmax(\mathbf{v}^{Global}), \tag{6}$$

where $\mathbf{v}^{Global} \in \mathbb{R}^l$ are parameters to be learned.

If we directly take the personalized weights $w_{u,p}$ as parameters to be learned, there will be $|\mathcal{U}| \times |\mathcal{P}|$ parameters to learn. However, training samples are usually insufficient for the weight learning. Following the idea in Section3.2, we assume the personalized weights are determined by some latent factors of users and meta-paths. So we can measure users' preferences on meta-paths with their embedding representations using an attention mechanism. Since users are entities and meta-paths are relations, their embeddings $\mathbf{e}^U$, $\mathbf{e}^P$ should be in different latent spaces [Fu *et al.*, 2017]. Therefore, we need to transform their embeddings into a same space, then take the dot product of transformed embeddings as the measure of users' preferences on meta-paths, then we define personal weight vector $\mathbf{w}_u = (w_{u,1}; \cdots ; w_{u,l})$ as follow:

$$\mathbf{w}_u = softmax((\mathbf{M}_{pa}\mathbf{E}^P)^\top \cdot \mathbf{M}_{ua}\mathbf{e}_u^U) \tag{7}$$

where $\mathbf{E}^P = (\mathbf{e}_1^P, \cdots , \mathbf{e}_l^P)$ is the embedding matrix of meta-paths, and $\mathbf{M}_{ua}, \mathbf{M}_{ua} \in \mathbb{R}^{d\times d}$ are the matrices for transforming users and meta-paths into the same attention space, respectively.

The HueRec model generates a recommendation list based on the predictions of users' interests on items. To measure the effectiveness of the preference learning on meta-paths and recommendation results, we define a loss function as follows.

$$Loss_{u,i}^{Rec} = -r_{u,i}log(\sigma(\widehat{r}_{u,i})) - (1 - r_{u,i})log(1 - \sigma(\widehat{r}_{u,i})) \tag{8}$$

where $r_{u,i}$ represents the implicit feedback of user u to item i. By minimizing the loss function in the Eq. 8, users' preferences on meta-paths can be learned. Embeddings of users, items and meta-paths can also be trained based on implicit feedback data.

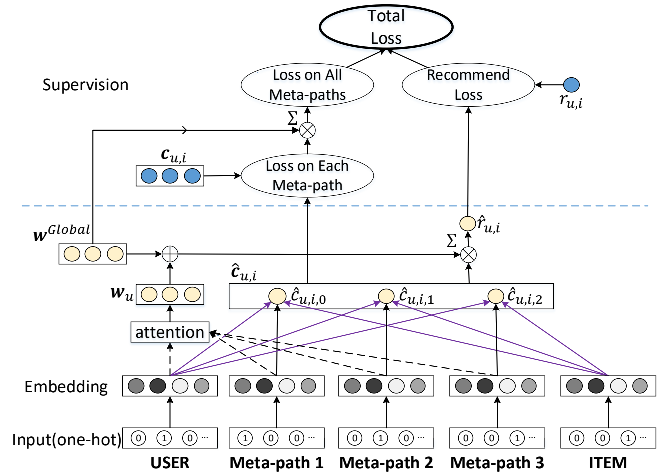

Figure 2: The training architecture of the HueRec. The part below the blue dashed line makes prediction and the above part describe the supervision process. The The $\mathbf{c}_{u,i}$ indicates the real interactions between the user u and the item i based on meta-paths, and the $\mathbf{r}_{u,i}$ is the implicit feedback.

3.4 Parameters Learning

To learn the parameters of the proposed model, we design an end-to-end training architecture, which is shown in the Fig. 2. The total objective is to minimize the following loss function:

$$Loss^{Total} = \frac{1}{|\mathcal{U}||\mathcal{I}|} \sum_{u\in\mathcal{U}} \sum_{i\in\mathcal{I}} (Loss_{u,i}^{Rec} + \lambda_1 Loss_{u,i}^{HIN}) + \lambda_2 \|\mathbf{\Theta}\|_2^2 \tag{9}$$

where λ_1 is the trade-off coefficient to control the importance of the $Loss^{HIN}$. We apply L2 regularization on all parameters $\mathbf{\Theta}$ to prevent the possible over-fitting and λ_2 is the regularization coefficient.

As the final loss function in Equation 9 is differentiable, we adopt gradient descent based algorithms to optimize it. In this paper, we appeal to Adaptive Moment Estimation (Adam) [Kingma and Ba, 2015] to minimize the total loss. It is worth noting that the global weights on meta-paths w_p^G are only determined by the recommendation loss $Loss^{Rec}$. As a result, we exclude the gradient from the $Loss^{HIN}$ when calculating the gradient of w_p^G.

Training all pairs in the Cartesian product of $\mathcal{U} \times \mathcal{I}$ for full gradient in each update step is not feasible. So we sample $\mathcal{X}$ in $\mathcal{U} \times \mathcal{I}$ with equal number of positive samples and negative samples. Then the computational complexity for one iteration training is $O(l \cdot d^2 \cdot |\mathcal{X}|)$ since Eq. 7 costs $O(l \cdot d^2)$ training time. Here we implement the HueRec model using the python library of TensorFlow[1].

4 Experiments

In this section, we first introduce the experimental datasets, evaluation metrics and experimental settings. Then we evaluate the effectiveness of HueRec compared with several state-of-the-art recommendation methods.

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

Dataset	Relations(A-B)	Number of A	Number of B	Number of (A-B)
Yelp	User-Business	14964	8954	47073
	User-User	4625	4625	26642
	Business-State	8165	13	8165
	Business-City	8165	154	8165
	Business-Category	8165	716	24848
	Business-Star	8165	9	8165
ML	User-Movie	2113	10109	855598
	Movie-Actor	10174	21185	52768
	Movie-Country	10180	70	10180
	Movie-Director	10155	4060	10155
	Movie-Genre	10197	20	20809
	Movie-Tag	7155	3800	23664
Douban	User-Movie	3022	6971	195493
	User-Group	2212	2269	7054
	User-Location	2491	244	2491
	Movie-Actor	5438	3004	15585
	Movie-Director	3014	789	3314
	Movie-Tag	6786	36	15598

Table 1: Statistics of datasets

4.1 Experimental Setup

Datasets

Three real-world data sets with the rich heterogeneous information is used as experimental data, including Yelp[2], MovieLens[3] and Douban Movie[4]. The Yelp dataset records users' ratings on local businesses and contains users' social relations and businesses' attribute information. The MovieLens dataset contains users' ratings on movies and some attribute information of movies. The Douban Movie dataset contains user attributes and movie attributes besides users' ratings. Detailed descriptions of the three datasets are shown in Table 1.

We select all meta-paths within three steps , since long meta-paths are likely to introduce more noise than useful information [Sun *et al.*, 2011].

Evaluation Methodology and Metrics

Since we focus on implicit feedback, we adopt the same processing as in [Yu *et al.*, 2014] to treat a rating record as a implicit (positive) feedback if the rating score exists.

Three-fold cross-validation is adopted to better evaluate the performances of different algorithms. The entire implicit feedback records of each dataset are randomly split into three roughly equally sized subsets. Each time, we use two-thirds of feedback records as the training set, the rest one-third as the testing set. The process is repeated three times with each of the three subsets used exactly once as the testing data. Experimental results are recorded as the average of the three runs.

For evaluation metrics, we adopt four widely used evaluation metrics for top-N recommendation [Wang and Blei, 2011; Van den Oord *et al.*, 2013; Hu *et al.*, 2018], including Precision at top k (Prec@k), Recall at top k (Recall@k), Mean Average Precision at top k (MAP@k) and Normalized Discounted Cumulative Gain at top k (NDCG@k). The former two are used to test model's ability of distinguishing items of users' interest, while the latter two focus on the position of preferred items in the ranked list.

[2]https://www.yelp.com/dataset_challenge

[3]https://grouplens.org/datasets/movielens/

[4]http://movie.douban.com

Baselines

We compare our model with these following methods:

-MF [Koren *et al.*, 2009]: Matrix factorization (MF) is a commonly used model-based collaborative filtering method.

-BPR [Rendle *et al.*, 2009]: Bayesian Personalized Ranking (BPR) is a typical pairwise learning method to rank for personalized recommendation based on implicit feedback.

-FM [Rendle, 2010]: Factorization machine (FM) models all interactions between each pair of features to estimate the target.

-HeteRec [Yu *et al.*, 2014]: HeteRec is a meta-path similarity based recommendation method. It first apply non-negative matrix factorization on each meta-path, then weighted ensemble all meta-path predicted scores.

-FMG$_{rank}$ [Zhao *et al.*, 2017]: FMG is a state-of-the-art HIN based recommendation model, which adopts a two-stage extraction (MF for HIN) and exploitation (FM for recommendation) process. We modify its optimization objective as pairwise ranking loss as in BPR for top-N recommendation.

-HueRec$_{noAtt}$: It is our model without attention mechanism as personalized meta-path weights.

-HueRec$_{noUni}$: It is a variant of our model in which we don't embed user and item into a shared latent space. It means we factorize meta-path connection matrix into a user embedding matrix and an item embedding matrix separately based on each meta-path.

4.2 Model Comparison

Experimental Design

We first randomly initialize model parameters with truncated Gaussian distribution with mean value 0 and standard deviation 0.02, then adopt Adaptive Moment Estimation (Adam) optimizer and L2 regularization coefficient λ_2 as 0.0001 for all models' optimization for a fair comparison. We set the dimension of embeddings to 64 for all models if needed. For our model, the coefficient λ_2 for HIN exploration part is set to 1. For baseline models, we tune their parameters based on authors' implementation if they exist, otherwise we tune to their best.

Experimental Results

Table 2 shows the performances of HueRec and baseline algorithms on different datasets with different evaluation metrics. The major findings from the experimental results are summarized as follows:

(1) Our HueRec model performs consistently better than all baselines on the three datasets according to different evaluation metrics. The results indicate the effectiveness of HueRec to extract HIN information and make top-N recommendation.

(2) Among side-information aware models, FM performs well on the Yelp dataset but fails on MovieLens and Douban Movie datasets. This may be due to the excessive attribute features in the latter two datasets. For example, MovieLens

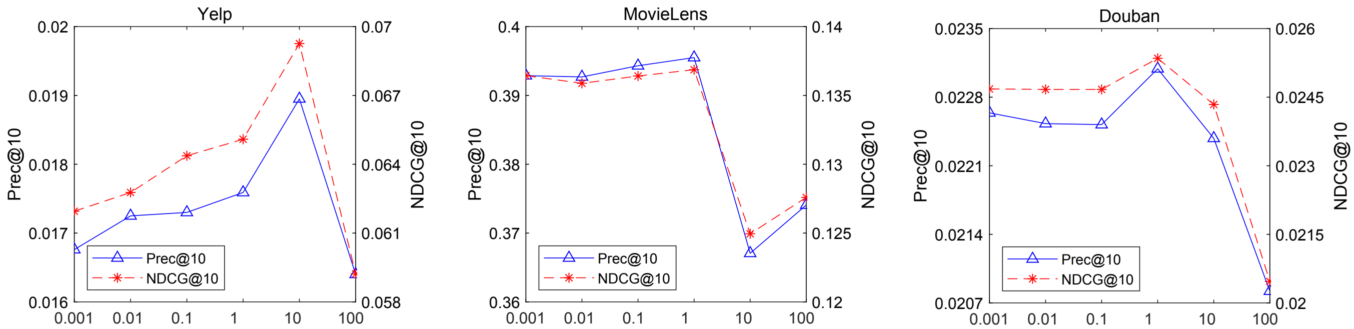

Figure 3: Performance of HueRec with different HIN exploration weight on three dataset

.

Dataset	Method	MAP@10	Prec@10	Recall@10	NDCG@10
Yelp	MF	0.0360	0.0148	0.0896	0.0537
	BPR	0.0337	0.0139	0.0901	0.0510
	FM	*0.0391*	*0.0167*	*0.1031*	*0.0595*
	HeteRec	0.0134	0.0062	0.0382	0.0211
	FMG_{rank}	0.0192	0.0089	0.0572	0.0307
	HueRec_{noAtt}	0.0422	0.0168	0.1121	0.0633
	HueRec_{noUni}	0.0380	0.0154	0.1019	0.0574
	HueRec	**0.0434**	**0.0176**	**0.1139**	**0.0651**
	Improve.	**11.00%**	**5.39%**	**10.48%**	**9.41%**
ML	MF	0.0336	0.3121	0.0613	0.1006
	BPR	0.0424	0.3248	0.0668	0.1163
	FM	0.0151	0.1995	0.0352	0.0567
	HeteRec	0.0310	0.2884	0.0552	0.0938
	FMG_{rank}	*0.0464*	*0.3616*	*0.0759*	*0.1261*
	HueRec_{noAtt}	0.0460	0.3535	0.0724	0.1238
	HueRec_{noUni}	0.0466	0.3643	0.0737	0.1248
	HueRec	**0.0529**	**0.3955**	**0.0793**	**0.1369**
	Improve.	**24.76%**	**21.77%**	**18.71%**	**17.71%**
Douban	MF	0.0069	0.0202	0.0207	0.0205
	BPR	0.0086	0.0224	*0.0265*	0.0243
	FM	0.0067	0.0197	0.0202	0.0199
	HeteRec	0.0073	0.0202	0.0221	0.0211
	FMG_{rank}	*0.0089*	*0.0225*	0.0259	*0.0244*
	HueRec_{noAtt}	0.0085	0.0224	0.0266	0.0243
	HueRec_{noUni}	0.0085	0.0224	0.0267	0.0242
	HueRec	**0.0092**	**0.0232**	**0.0272**	**0.0254**
	Improve.	**6.98%**	**3.57%**	**2.64%**	**4.53%**

Table 2: Performance comparisons of different methods.

has 10k movies while over 20k actors, which imports too much noise. While HIN based methods (HeteRec, FMG, HueRec) performs more stably on the latter two datasets, because they only focus on attribute sharing relations instead of what these attributes exactly are. This demonstrates the advantage of HIN based models.

(3) As a two step model, HeteRec and FMG perform poorly on the Yelp dataset. This may be caused by the sparsity of the Yelp dataset. Some errors may occur in the information extraction step, which may mislead the following recommendation model.

(4) Our complete model HueRec performs consistently better than the two variants HueRec_{noAtt} and HueRec_{noUni}. This confirms that it is essential to embed users and items into a unified latent space based on all meta-paths, and to personalize the weights on meta-paths.

4.3 Hyper-Parameter Study

The weight of HIN exploration part λ_1 is an important parameter for the proposed model HueRec. It determines the relative importance of information extracted from HINs for user and item modeling. When λ_1 is large, we model users and items mainly based on structural features of HINs. Otherwise, we model users and items mainly based on historical behaviors. Fig. 3 shows the performances of HueRec with different λ_1. When the weight is small (less than 10 for Yelp or 1 for MovieLens and Douban Movies), performances become better as the increasing of the weight. It indicates that extracting information from HIN is helpful to recommendation. When the weight continues to increase, performances become worse, which means not only useful information but also some noises have been extracted. To make a better trade-off for the model, we suggest to set the HIN exploration weight as 1.

5 Conclusion

In this paper, we propose an HIN based unified embedding model for recommendation,*i.e.* HueRec. We design a method to learn unified embeddings, which can extract and fuse information from different meta-paths. We use an attention mechanism to combine predicted scores of different meta-paths. Furtherly, we conduct information extraction and recommendation simultaneously to alleviate information loss. Extensive experiments on three real datasets demonstrated the effectiveness of HueRec.

As future work, we would like to design a more sophisticated model, in place of the weighted linear sum in the recommendation part, to further improve the recommendation performance.

Acknowledgements

This work was partially sponsored by National Natural Science Foundation of China (Grant No. 61672062), National 863 Program of China (Grant No. 2015AA016009) and Delta Innovation Research Program.

References

[Bahdanau *et al.*, 2015] Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations*, 2015.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pages 2787–2795, 2013.

[Fu *et al.*, 2017] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. HIN2Vec: Explore meta-paths in heterogeneous information networks for representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1797–1806, 2017.

[Han *et al.*, 2018] Xiaotian Han, Chuan Shi, Senzhang Wang, Philip S. Yu, and Li Song. Aspect-level deep collaborative filtering via heterogeneous information networks. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3393–3399, 2018.

[Hu *et al.*, 2018] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S. Yu. Leveraging meta-path based context for top- n recommendation with a neural co-attention model. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1531–1540, 2018.

[Jiang *et al.*, 2018] Zhengshen Jiang, Hongzhi Liu, Bin Fu, Zhonghai Wu, and Tao Zhang. Recommendation in heterogeneous information networks based on generalized random walk model and bayesian personalized ranking. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, pages 288–296, 2018.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2015.

[Koren *et al.*, 2009] Yehuda Koren, Robert M Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[Liu *et al.*, 2018] Hongzhi Liu, Zhonghai Wu, and Xing Zhang. CPLR: Collaborative pairwise learning to rank for personalized recommendation. *Knowledge-Based Systems*, 148:31–40, 2018.

[Liu *et al.*, 2019] Hongzhi Liu, Zhengshen Jiang, Yang Song, Tao Zhang, and Zhonghai Wu. User preference modeling based on meta paths and diversity regularization in heterogeneous information networks. *Knowledge-Based Systems*, 2019.

[Rendle *et al.*, 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.

[Rendle, 2010] Steffen Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000, 2010.

[Shi *et al.*, 2015] Chuan Shi, Zhiqiang Zhang, Ping Luo, Philip S. Yu, Yading Yue, and Bin Wu. Semantic path based personalized recommendation on weighted heterogeneous information networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 453–462, 2015.

[Shi *et al.*, 2016] Chuan Shi, Jian Liu, Fuzhen Zhuang, Philip S Yu, and Bin Wu. Integrating heterogeneous information via flexible regularization framework for recommendation. *Knowledge and Information Systems*, 49(3):835–859, 2016.

[Shi *et al.*, 2017] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and Philip S Yu. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 29(1):17–37, 2017.

[Shi *et al.*, 2019] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and S Yu Philip. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 31(2):357–370, 2019.

[Sun *et al.*, 2011] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In *Proceedings of the VLDB Endowment*, 4(11):992–1003, 2011.

[Van den Oord *et al.*, 2013] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In *Advances in neural information processing systems*, pages 2643–2651, 2013.

[Wang and Blei, 2011] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 448–456, 2011.

[Yu *et al.*, 2013] Xiao Yu, Xiang Ren, Quanquan Gu, Yizhou Sun, and Jiawei Han. Collaborative filtering with entity similarity regularization in heterogeneous information networks. In *2nd IJCAI Workshop on Heterogeneous Information Network Analysis*, 2013.

[Yu *et al.*, 2014] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 283–292, 2014.

[Zhao *et al.*, 2017] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. Meta-graph based recommendation fusion over heterogeneous information networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 635–644, 2017.