Recommendation Based Heterogeneous Information Network and Neural Network Model

Cong Zhao[1], Yan Wen[1](✉), Ming Chen[2], and Geng Chen[3]

[1] College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China
wenyan84@hotmail.com
[2] State Grid Shandong Electric Power Company,
Qingdao Power Supply Company, Qingdao 266500, China
[3] College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Abstract. With the advent of the Internet era, the recommendation system has developed rapidly. Heterogeneous information networks representation learning is widely used in recommendation systems due to its advantages in complex information modeling. Although the performance of the recommendation system has been improved, there are still two shortcomings: 1. There is a lot of noise data in the instances of the meta-path generated by random walks of meta-paths, which will reduce the performance of the recommendation system. 2. Traditional recommendation algorithms fail to make full use of the relevant meta-path information in heterogeneous information networks, which makes the recommendation results lack of interpretability. To solve these problems, we propose a recommendation system based on heterogeneous network representation learning and neural network model. Firstly, use the matrix factorization and the similarity calculation to select the meta-path instances with good quality as the pre-training vectors of the recommendation system. Then, we combine LSTM with Attention mechanism to learn the user, item and meta-path embeddings, and use MLP to make prediction after fusion to jointly improve the recommendation effect. We conducted experiments on Movielens datasets to evaluate the performance of our proposed recommendation.

Keywords: Heterogeneous information network representation learning · Neural network · Recommendation system

1 Introduction

The rapid development of the Internet era has brought the huge amount of information resources, which makes users can't accurately obtain the information they need. Therefore, as a kind of information retrieval tool in the Internet era, the recommendation system has developed rapidly.

Q. Wu et al. (Eds.): WiSATS 2020, LNICST 358, pp. 588–598, 2021.
https://doi.org/10.1007/978-3-030-69072-4_48

Currently, heterogeneous information networks (HIN) [1] are widely used in recommendation systems due to their advantages in complex information modeling. In Fig. 1, we can see that the HIN contains multiple types of entities connected by different types of relations. Meanwhile, network representation learning [2], which aim at learning the low-dimensional representation vectors of each node has shown certain potential in structural feature extraction. Heterogeneous information network representation learning is an extension of the homogeneous network representation learning, which emphasizes both on the complex structure and rich semantics, and has shown effectiveness in many recommendation tasks.

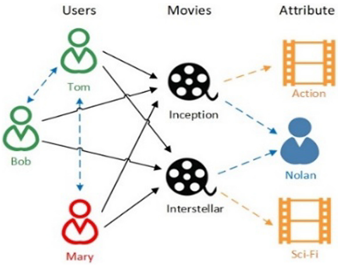


Fig. 1. Movie heterogeneous information network sample diagram

Although existing recommendation methods based on heterogeneous network representation learning have achieved performance improvement to some extent, they still face some challenges: (1) In most existing HIN representation learning works, there is a lot of noise data in the instances of the meta-path generated by random walks of meta-paths, which will reduce the performance of the recommendation system. (2) Most existing the recommendation algorithms only model the interaction between users and items, and fails to make full use of the meta-paths in HIN with abundant auxiliary information, which makes the recommendation results lack of interpretability.

In order to solve the above problems, we propose a recommendation system based on heterogeneous network representation learning and neural network model. The main contributions of this paper are as follows: (1) In order to filter out noisy meta-path instances, the matrix factorization algorithm is used to generate the latent factors for users and items, which are future used to calculate the similarity between nodes guide the meta-path based random walk, calculate the cosine similarity between nodes, and Top-k meta-path instances with high score are selected. (2) We propose a neural network recommendation framework that combines LSTM with Attention mechanism. This framework can learn the three feature representations of user, item and meta-path information, and make predictions after the integration of the three to improve the recommendation effect. (3) We conducted experiments on Movielens dataset to evaluate the performance of our methods. The results show that we prove the good performance of the proposed model in the recommended task.

2 Related Work

In the early work on recommendation system, the classical recommendation algorithm matrix factorization, is to model users' and items' preferences by learning two latent semantic matrices which are used as a prediction function for the recommended items. However, the matrix factorization algorithm will generate cold start problem, we try to improve the effect of recommendation by adding auxiliary information. For example, Ma et al. [1] proposed adding auxiliary data such as social relations to matrix decomposition to recommend users. The model proposed by Ling et al. [4] not only considers scoring information, but also adds comment information to it, realizing the combination of content-based filtering and collaborative filtering for prediction.

In recent years, representation learning has become a hot research. Among them, representative learning algorithms based on heterogeneous information network include metapath2vec [5], HIN2vec [6], HERec [7], etc. At present, most online services are typical heterogeneous information networks which can make full use of various rich semantic information to make personalized recommendations for users. Chuan Shi et al. [7] proposed HERec algorithm, which through representation learning, obtain vector representations of users and commodities and fuse them, use matrix factorization algorithm to complete the prediction. In the MCRec proposed by Binbin Hu et al. [8], Using PCRW sampling sequence, the embedded representation of different meta-paths is obtained through the convolutional neural network, using attention mechanisms to fuse embedded representations, then, a layer of fully connected neural network is used to predict the user's rating of the item.

3 Preliminary

Definition 1. Heterogeneous information network [1]. A HIN is denoted as G = (V, E), node set V and edge set E. It is also associated with an object type mapping function $\varphi : V \rightarrow A$ and a link type mapping function $\delta : E \rightarrow R$. A and R denote the sets of predefined object and link types, where $|A| + |R| > 2$.

Definition 2. The network schema is denoted as S = (A; R). It is a meta template for an information network G = (V, E) with the object type mapping $\varphi : V \rightarrow A$ and the link type mapping $\delta : E \rightarrow R$, which is a directed graph defined over object types A, with edges as relations from R.

Definition 3. Meta-path [9]. The meta-path ρ defined as $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_t} A_{t+1}$ (abbreviated as $A_1 A_2 \cdots A_{t+1}$) on the network schema S $=$ (A, R), represents a compound relationship between node type A_1 and A_{t+1} through node type sequence $A_1, A_2, \ldots A_t$, and $R = R_1 \mathrm{o}\, R_2 \mathrm{o} \cdots \mathrm{o}\, R_t$ represents the compound operator on the relationship. Some meta-paths with difference semantics are defined as in Table 1.

4 The Proposed Model

4.1 Model Overview

The main framework of this paper is the combination of heterogeneous information network representation learning and neural network. First, matrix factorization algorithm

Table 1. Examples of meta-path

Meta-path	Meaning	Examples
U-U (User-User)	Friend	Tom-Mary, Tom-Bob
U-M-U (User-Movie-User)	Users who watch the same movie	Tom-Avater-Bob
U-M-D-M-U (User-Movie-Director-Movie-User)	Users who watch movies made by the same director	Tom-Avatar-Cameron-The Titanic-Bob

is used to evaluate the meta-path instances generated by meta-path-based random walks. For each meta-path, the instances are assessed by average similarity between adjacent nodes along the instance path, and TOP-K instances are selected as pre-training vectors for the recommendation system. Then a neural network recommendation framework combining LSTM and Attention mechanism (in Fig. 2) is proposed. This framework can learn the feature representations of user, item and meta-paths, and make prediction by using MLP, so as to jointly improve the recommending performance.

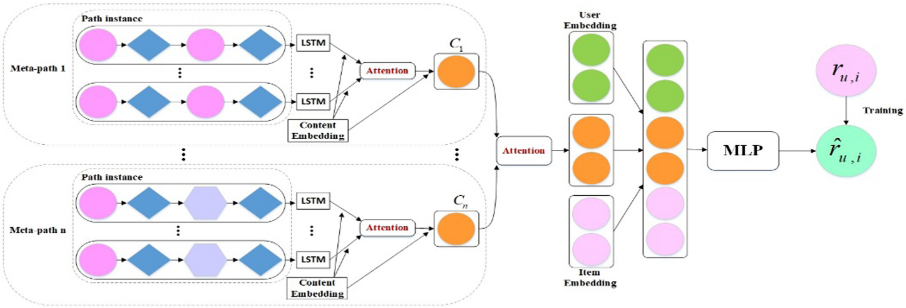


Fig. 2. Overall framework

4.2 Selection of Meta-path Instances

A meta-path instance is a sequence of nodes generated by random walks a meta-path. The average similarity of the edges in the meta-path instance can be taken as its score, then the Top-k instances are selected as the input data.

Matrix Factorization on Edge Type. For a give edge type R = <A1, A2 >, all the edges with the type R is, the corresponding matrix is defined as M, the element $r_{u,i}$ means the value of u^{th} row and i^{th} column in M, and is set to 1 when there is an edge between u and i, and 0 otherwise. The matrix M can be factorized into two latent matrix X and Y as M $\approx X^T \cdot Y$, where X = [x1,x2,…,xn] and Y = [y1, y2,…,ym] represent latent matrices for node type A1 and A2 under edge type R respectively, and xi and yj represent the latent vectors. The latent vector for each node can be obtained by optimizing formula (1).

$$\min_{U,L} \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \| W_{i,j} - U_{i.}^{T} L_{.j} \|_F^2 + \frac{\lambda}{2} \left(\| U \|_F^2 + \| L \|_F^2 \right) \tag{1}$$

Meta-path-based random walks on heterogeneous networks is based on following transition probability:

$$P(n_{t+1} = x | n_t = v, \rho) = \begin{cases} \frac{1}{\left| N^{A_{t+1}}(v) \right|}, (v, x) \in \varepsilon \ and \ \emptyset(x) = A_{t+1}; \\ 0, otherwise, \end{cases} \tag{2}$$

Where n_{t+1} is the next node, n_t is the current node, v is the node type, ρ is the meta-path, and the denominator $\left| N^{A_{t+1}}(v) \right|$ represents the number of candidate nodes according to the meta-path.

Similarity Measure for Node Pairs on a Mate Path Instance. Suppose the meta-path ρ is *umum*, and one of its instances is $node^1, node^2, node^3 \ldots\ldots, node^k$, and each node can be represented by an n-dimensional vector. The similarity between adjacent nodes in the meta-path instance is based on the cosine similarity measure shown in Eq. (3).

$$sim\left(node^1, node^2 \right) = \frac{\sum_{i=1}^{n} node_i^1 \times node_i^2}{\sqrt{\sum_{i=1}^{n} \left(node_i^1 \right)} \times \sqrt{\sum_{i=1}^{n} \left(node_i^2 \right)}} \tag{3}$$

Meta-path Instance Score. The score of a meta-path instance ρ is the average of the similarity value of the node pairs along the instance as shown in (4) where k is the number of nodes on σ.

$$score_\sigma = \frac{\sum_{j=1}^{k-1} sim\left(node^j, node^{j+1} \right)}{k - 1} \tag{4}$$

The higher the score is, the higher the similarity between the nodes in the meta-path instance is, which makes the semantics of the meta-path instance more explicit.

4.3 User and Item Embedding

As look up can simplify the operation and low time complexity, we use look up to produce the user embedding and item embedding. Formally, given a user-item pair < user, item> , let a_u and b_i denote their one-hot representations.

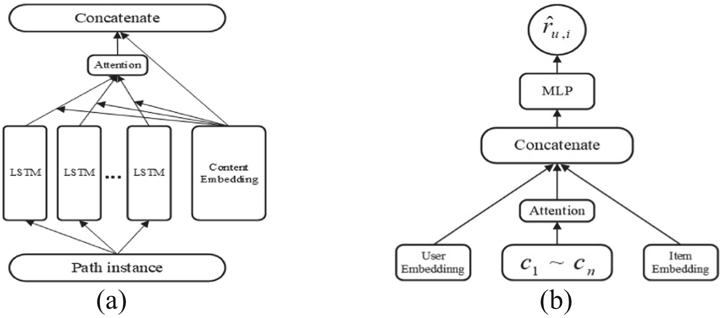
The lookup layers correspond to two parameter matrices A and B, which store the latent factors for users and items respectively. And d is the dimension size of user and item embeddings, and U and I are the total number of users and items respectively. The lookup operation is implemented as follows:

$$x_u = A^T \cdot a_u \tag{5}$$

$$y_i = B^T \cdot b_i \tag{6}$$

4.4 Meta-path Embedding

In order to capture particular semantics for different meta-paths, each meta-path is processed separately. As LSTM is good at sequence modeling and the Attention mechanism can obtain the importance of each meta-path instance. We combine the LSTM and the Attention mechanism to produce the meta-path instances embedding, content embedding and path instances' weight distribution, then, concatenate path instance embedding obtained by the Attention mechanism and content embedding to obtain a single meta-path embedding (in Fig. 3(a)). Similarly, different meta-paths play different roles for determining the nodes relevance, so we introduce the Attention mechanism to put different weight on each meta-path embedding. Each individual meta-path embedding is fused into one final meta-path representation (in Fig. 3(b)).



(a) (b)

Fig. 3. Model processing

Use the Combination of LSTM and Attention Mechanism to Obtain the Embedding of a Single Meta-path. While extracting each meta-path instance embedding, we also trained content embedding. When all path instances under this meta-path are trained, content embedding includes the representation of the content of this meta-path.

We use LSTM to learn the sequence embedding of path instance i, where X^i is the node embedding matrix, and θ represents all relevant parameters in LSTM, as follows:

$$h_i = LSTM\left(X^i; \theta\right) \tag{7}$$

Then, through the Attention mechanism to put different weight on each meta-path instance embedding. Where, h_i is the sequence embedding of the meta-path instance obtained by LSTM, u_i is the feature matrix after nonlinear transformation, u_p represents the content embedding, it is continuously updated as the model training, α_i represents the Attention mechanism distribution coefficient of different meta-path instances, o is the output result of Attention mechanism,

$$u_i = tanh(W_s h_i + b_s) \tag{8}$$

$$\alpha_i = \frac{\exp\left(u_i^T u_p\right)}{\sum_i \exp\left(u_i^T u_p\right)} \tag{9}$$

$$o = \sum_{i} \alpha_i h_i \tag{10}$$

We concatenate the Attention mechanisms' result o and the content embedding u_p together to form a single meta-path embedding p_j, as follows:

$$p_j = concatenate\left(o, u_p\right) \tag{11}$$

Obtained Different Weight of Different Meta-paths Embedding. Different metapaths play different roles for determining the nodes relevance. In Fig. 4, we use the Attention mechanism to obtain the weight of different meta-paths ($C_1 \sim C_n$).

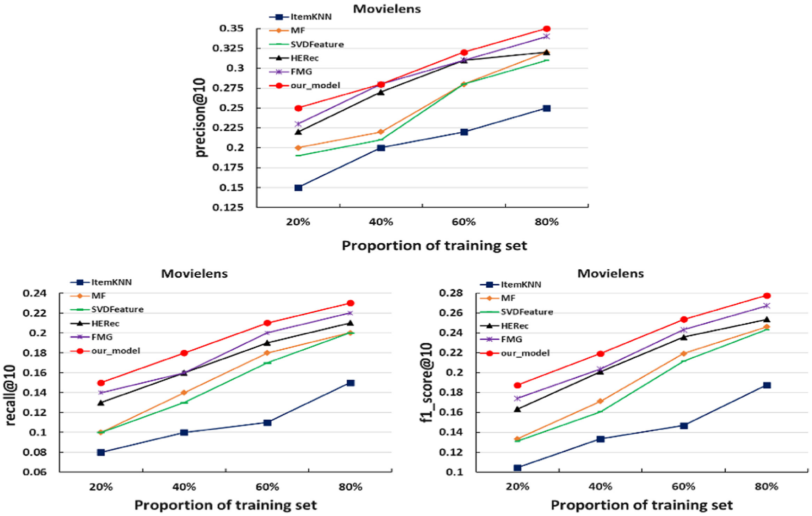


Fig. 4. The precision and recall on the Movielens dataset

$$v_j = tanh\left(W_k p_j + b_k\right) \tag{12}$$

$$\beta_j = \frac{\exp\left(v_j^T u_m\right)}{\sum_j \exp\left(v_j^T u_m\right)} \tag{13}$$

$$v = \sum\nolimits_j \alpha_j p_j \tag{14}$$

Where, p_j represents the embedding of each meta-path, and β_j represents the Attention mechanism distribution coefficient of different meta-path. v is the result of weight of different meta-paths embedding.

4.5 Interaction of User, Item and Meta-path Embedding

In Fig. 4, we concatenate the meta-path embeddings with user embedding and item embedding, as follows:

$$out = concatenate(x_u, v, x_i) \tag{15}$$

4.6 Model Prediction

In Fig. 4, we feed the *out* to the MLP component, in order to model the complex non-linear interaction among the three embeddings.

$$\hat{\gamma}_{\mathrm{u,i}} = MLP\,(\mathrm{out}) \tag{16}$$

Defining objective functions for model optimization is the key to learning a good recommendation model. In our task, the objective for an interaction <u, i> can be formulated as follows:

$$l_{u,i} = -log\,\hat{\gamma}_{u,i} - E_{j \sim P_{neg}}\left[\log\left(1 - \hat{\gamma}_{u,j}\right)\right] \tag{17}$$

Where the first term models the observed interactions and the second term models the negative feedback from the noise distribution P_{neg}.

5 Experiment

We conducted experiments on Movielens datasets and proved the effectiveness of our proposed method by comparing it with other methods.

5.1 Dataset

We use Movielens datasets, which contains data for users and their ratings on movies, including 943 users and 1,682 movies.

The meta-paths used are shown in Table 2. The user and item have an embedded dimension of 128, a random walk path length of 10, and a window size of 3. The learning rate is set to 0.001, the Gaussian distribution is used to randomly initialize the model parameters, the batch size is 128, the optimizer uses Adaptive Moment Estimation (Adam), and the number of iterations is set to 100. When the loss has not dropped for 20 consecutive rounds, stop training.

Table 2. Meta-paths selected for each dataset

Dataset	Meta-paths
Movielens	UMU, UMDMU, UMAMU, UMTMU, UUM

We consider comparing with the following methods:

- ItemKNN [10]: A classic collaborative filtering method that recommends similar items based on previous items.
- MF [11]: Standard matrix factorization method, which uses the cross-entropy loss function to modify its optimization loss.
- SVD Feature$_{mp}$: A variant of SVDFeature [12], which uses the HIN embedding method meta-path2vec ++ to extract user and project embedding.
- FMGrank: Based on FMG [13], the optimization objective was modified to the paired ranking loss in BPR for scoring prediction.
- HERec [7]: A recommendation method based on heterogeneous networks embeddings, which is fused with matrix methods for recommendation.

5.2 Overall Performance

We divide the entire dataset into a training set and a test set. We use different proportions of data (20%, 40%, 60%, 80%) as the training set. We use ranking-based accuracy (Prec @ K), recall (Recall @ K) and f1 value (F1-score @ K) as to the metrics to evaluate the recommended performance of our model. In Fig. 4, The result shows that: (1) Our proposed method is always superior to all comparison models. Compared with other HIN-based methods, our model uses a method based on heterogeneous network representation learning combined with LSTM and Attention mechanism to fuse the three aspects of user, item and meta-path information for recommendation tasks, making recommendation result more accurate. (2) The model's performance improves as the training data increases.

5.3 Meta-path Instances Selection

In order to verify the effectiveness of the meta-path instance selection method, we compare the performance (F1-score) between model with filtered and unfiltered meta-path instances as data input, as shown in Fig. 5. The results show that by filtering out the noisy instances, the recommendation performance can be effectively improved. the recommended model.

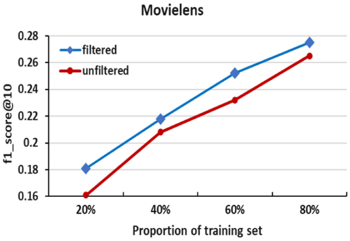


Fig. 5. Meta-path instance selection

5.4 Meta-path Attention Weight Visualization

The attention weights for different meta-paths is visualized in Fig. 6, which shows the impact of different meta-paths on recommendation performance. We can find that for the Movielens dataset, the “UMDMU” meta-path takes more weight than other meta-paths. Therefore, we believe that some meta-paths can improve the performance of the model more than others.

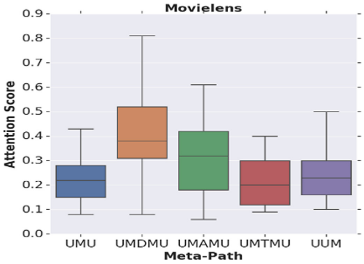


Fig. 6. Meta-path attention weight visualization

6 Summary

In this paper, we propose a recommendation system based on heterogeneous network representation learning and neural network model. We selected meta-path instances with better quality as the pre-training vectors of the recommendation system. We combine the LSTM and Attention mechanism to learn the three aspects of user, item, meta-path information, use the MLP to predict, so as to improve the recommendation effect. In the experiment, we prove that our model improves the performance of the recommendation system. By concatenate with users, items and meta-path embedding, the embedded vector information is enriched, which not only improves the recommendation effect, but also makes full use of the meta-path information to enhance the interpretability of the recommendation results.

Acknowledgements. This work was supported by the Qingdao Philosophy and Social Sciences Planning Project (QDSKL1801131), the 2018 Postgraduate Tutors’ Guidance Ability Improvement Project of Shandong Province (84), the National Natural Science Foundation of China under Grant No. 61701284, the Innovative Research Foundation of Qingdao under Grant No.19-6-2-1-cg.

References

1. Sun, Y., Han, J.: Mining heterogeneous information networks: principles and methodologies. Synth. Lect. Data Min. Knowl. Discov. **3**, 1–159 (2012)
2. Chen, H., Perozzi, B., Al-Rfou, R., et al.: A Tutorial on Network Embeddings (2018)

3. Ma, H., Zhou, D., Liu, C., et al.: Recommender systems with social regularization. In: Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, 9–12 Feb 2011. DBLP (2011)
4. Ling, G., Lyu, M.R., King, I.: Ratings meet reviews, a combined approach to recommend. In: Proceedings of the 8th ACM Conference on Recommender Systems, pp. 105–112 (2014)
5. Dong, Y., Chawla, N.V., Swami, A.: metapath2vec: scalable representation learning for heterogeneous networks. In: ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM (2017)
6. Fu, T.Y., Lee, W.C., Lei, Z.: HIN2Vec: explore meta-paths in heterogeneous information networks for representation learning. In: Proceedings of the 26th ACM International Conference on Information and Knowledge Management, pp. 1797–1806. ACM Press (2017)
7. Shi, C., Hu, B., Zhao, W.X., et al.: Heterogeneous information network embedding for recommendation. IEEE Trans. Knowl. Data Eng. **31**(2), 357–370 (2019)
8. Hu, B., Shi, C., Zhao, W.X., et al.: 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, pp. 1531–1540. ACM (2018)
9. Sun, Y., Han, J., Yan, X., et al.: Pathsim: meta path-based top-k similarity search in heterogeneous information networks. Proc. VLDB Endow. **4**(11), 992–1003 (2011)
10. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web, pp. 285–295 (2001)
11. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**, 30–37 (2009)
12. Chen, Tianqi., Zhang, Weinan., Qiuxia, Lu., Chen, Kailong., Zheng, Zhao, Yong, Yu.: SVD-Feature: a toolkit for feature-based collaborative filtering. J. Mach. Learn. Res. **13**(2012), 3619–3622 (2012)
13. Zhao, H., Yao, Q., Li, J., Song, Y., Lee, D.L.: Metagraph based recommendation fusion over heterogeneous information networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 635–644 (2017)