Representation Learning with Depth and Breadth for Recommendation using Multi-view Data

Xiaotian Han[1], Chuan Shi[1(✉)], Lei Zheng[2], Philip S. Yu[2], Jianxin Li[3], and Yuanfu Lu[1]

[1] Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, China
[2] University of Illinois at Chicago, Chicago, USA
[3] The University of Western Australia
{hanxiaotian.h,luyfroot}@gmail.com,shichuan@bupt.edu.cn,{psyu,lzheng21}@uic.edu,jianxin.li@uwa.edu.au

Abstract. Recommender system has been well investigated in the past years. However, the typical representative *CF*-like models often give recommendation with low accuracy when the interaction information between users and items are sparse. To address the practical issue, in this paper we develop a novel **R**epresentation **L**earning with **D**epth and **B**readth (*RLDB*) model for better recommendation Specifically, we design a heterogeneous network embedding method and convolutional neural network based method to learn feature representations of users and items from user-item interaction structure and review texts, respectively. Furthermore, an end-to-end breadth learning model is proposed through employing multi-view machine technique to learn features and fuse these diverse types of features in a uniform framework. Extensive experiments clearly demonstrates that our model outperforms all the other methods in these datasets.

Keywords: Recommender System, Rating Prediction, Multi-view Machine, Heterogeneous Information Network Embedding

1 Introduction

In the past decade, the rapid growth of e-commerce is changing people's daily life such as online shopping, reading articles, and watching movies. Therefore, effective recommender systems are highly desirable to help the customers by presenting the products or services that are likely of interest to them. *Collaborative Filtering* (*CF*) is the most representative model which assumes that people who share similar preferences in the past tend to have similar choices in the future. *CF* model based recommendations is often lack of sufficient interaction between users and items.

Fortunately, there are lots of additional available data information that can help to alleviate the issue of insufficient interaction information between users and items, e.g., the review text. There are some studies [6, 5, 9] that have used review text to improve the recommendation results. McAuley et al in [6] and Ling et al in [5] still mainly used the structure information to construct user-item matrix and then utilized the review text as the auxiliary information to slightly adjust the user-item weights during matrix factorization. In contrast, Zheng et al [9] only used the review text to learn user behavior and item properties, which are then used to learn the hidden latent features for users and items jointly. But they ignored the effect of structure information in use. In addition, in most cases we need to treat the different views of data to be equally important as much as possible while the importance of noisy features should be discriminated automatically and effectively.

To do this, in this paper, we design a novel *R*epresentation *L*earning model with *D*epth and *B*readth (called *RLDB*) to effectively utilize two views of data (i.e., structure and text information) through two deep neural networks extracting feature representation and a breadth learning model fusing features.

The contributions of this paper are summarized as follows:

- We propose a novel HIN embedding method to extract structure information of users and items for recommendation. As well as we design a CNN and meta-path based heterogeneous network embedding method to effectively learn the representation of the structure information.
- We propose a end-to-end based breadth learning method through employing multi-view machine to effectively combine the structure representations and text representation together.
- We conduct extensive experiments on three real-world datasets to evaluate the effectiveness of RLDB.

2 Proposed Model

In order to integrate different views of data for improving recommendation, we develop a novel model - **R**epresentation **L**earning with **D**epth and **B**readth (*RLDB*), to model the representations of users, the representations of items, and their combinations.

2.1 Model Architecture

The overall architecture of the RLDB model is shown in Fig. 1. The proposed model can learn representations from different views via deep neural networks and effectively couple them with a multi-view machine for recommendation. In this paper, we take the structure information of rating network and text information in reviews of users and items as the four views. The details of RLDB are described below.

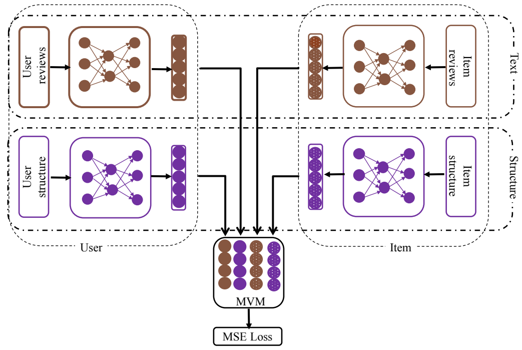


Fig. 1: The architecture of the proposed model

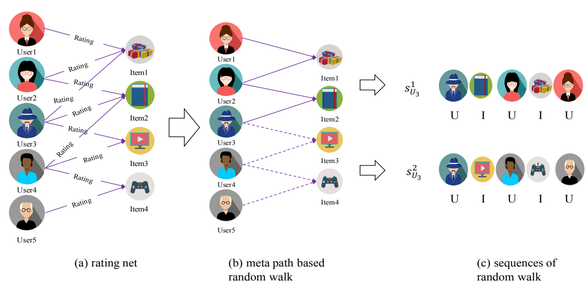


Fig. 2: The generation process of random walk sequences

2.2 Structure Representation

To take users' consumption habits into account, we construct an rating network as a type of heterogeneous information networks [8]. Then we model the representation of the interaction network structure of user/item via a neural network.

Meta-Path Based Random Walk In our model, we use meta-path based random walk to model the representation of the users' structure of rating network as well as the items'. In order to illustrate the meta-path based random walk, we show a toy example of the rating network to illustrate structure of the rating network in Fig. 2(a).

In order to illustrate the generation process, Fig. 2(b) shows the sequence generation process of the user U_3. We set L as 5. The nodes connected by the solid line make up the first sequence $s^1_{u_3}$ on behalf of the user U_3, which contains $U_3I_2U_2I_1U_1$. The dotted line connects the nodes to form the anther sequence $s^2_{u_3}$ containing $U_3I_3U_4I_4U_5$. Fig. 2(c) shows the result of the meta-path based random walk. The sequence $s^k_{U_j}$ represents the kth random walk starting with the node s_{U_j} on behalf of the user U_j. The $s^1_{U_1}$ in Fig. 2 is based on the mate-path $UIUIU$. After obtaining the sequence, we concatenate all the random walk sequences starting with s_{U_j} to represent the rating network structure of the user s_{U_j}.

$$S_{U_i} = s_{U_i}^1 \oplus s_{U_i}^2 \oplus s_{U_i}^3 \oplus \cdots \oplus s_{U_i}^n \tag{1}$$

Similar to the user, the concatenated sequence S_{I_i} on behalf of the item I_i's structure.

Learning the Representation of Structure After getting the concatenated sequences of random walks on behalf of the users or items, we learn the representation of the users and items rating network structure. First, we embed the users and items into a uniform vector space and update the embedding vector via the training process. the nodes are represented as a vector, so we represent this sequence as a two-dimensional matrix and feed it to a convolution neural network with several filters.

This feature map of the kth filter within the convolutional layer with the is as follow:

$$\hat{r}_{U_i}^k = \left[r_k^1, r_k^2, \cdots, r_k^3, \cdots, r_k^{n-h+1}\right] \tag{2}$$

where k represents the kth filter in the convolution layer.

We concatenate multiple features obtained by multiple filters (with varying window sizes) to generate the representation $R_{U_i}^s$ of the rating network structure on behalf of user U_i:

$$R_{U_i}^s = r_{U_i}^1 \oplus r_{U_i}^2 \oplus r_{U_i}^3 \oplus \cdots \oplus r_{U_i}^n \tag{3}$$

We can obtain the item I_i's representation $R_{I_i}^s$ of rating network structure in the same way.

As discussed before, the semantics underneath different paths are different. For example, in the conventional layer, the window size of 3 can model the semantics of the mate-path UIU while window size of 5 can capture the semantics of the mate-path $UIUIU$. Obviously, the distinct semantics under different paths will lead to different information.

2.3 Text Representation

In the part of text information, we utilize CNN to facilitate the deeper understanding of users' review texts. In the first layer, all the review texts for users or items are represented as matrices of word embedding [7] to capture the semantic information. The next layers are the common layers used in CNN based models to discover multiple features for users and items with the multiple features of different filters. The top layer is a max pooling layer which can find the most important features.

Consider the user review texts as an example. Let us denote all the reviews of a user U_1 consisting of n reviews as $d_{U_1I_1}, d_{U_1I_2}, d_{U_1I_3}, \cdots, d_{U_1I_i}, \cdots, d_{U_1I_n}$.

$$D_{U_i} = d_{U_iI_1} \oplus d_{U_iI_2} \oplus d_{U_iI_3} \oplus \cdots \oplus d_{U_iI_n} \tag{4}$$

where $d_{U_i I_j}$ indicates the review from user U_i to I_j and $\oplus$ is the concatenation operator.

This feature map of the kth filter within the convolutional layer with the is as follow:

$$\hat{r}_{U_i}^{k} = \left[r_k^1, r_k^2, \cdots, r_k^3, \cdots, r_k^{n-h+1}\right] \tag{5}$$

To capture the most important feature, we feed the output of a convolutional filter to a max pooling layer. Denote the $r_{U_i}^{k} = max\{\hat{r}_{U_i}^{k}\}$ as the feature corresponding to the kth filter. We concatenate multiple features obtained by multiple filters (with varying window sizes) to generate the representation $R_{U_i}^{t}$ of the review texts on behalf of user U_i :

$$R_{U_i}^{t} = r_{U_i}^{1} \oplus r_{U_i}^{2} \oplus r_{U_i}^{3} \oplus \cdots \oplus r_{U_i}^{n} \tag{6}$$

We can obtain the item I_i representation $R_{I_i}^{t}$ of review texts in the same way.

2.4 Fusion with Multi-view Machine

We obtain four representations of the rating network structure and review texts for users and items respectively.

Inorder to combine the four representations effectively, We further explore all vectors of these representation interactions up to the mth-order between inputs from m-view data by the multi-view machine [2]. Equation 7 shows the principle of multi-view machine that interacts all features to the mth-order between inputs from m views.

$$\hat{y} = \sum_{i_1=1}^{I_1+1} \cdots \sum_{i_m=1}^{I_m+1} w_{i_1,\ldots,i_m} \left(\prod_{v=1}^{m} z_{i_v}^{(v)} \right) \tag{7}$$

where the $\mathbf{z}^{(v)^T} = \left(\mathbf{x}^{(v)^T}, 1\right) \in \mathbb{R}^{I_v+1}, \forall v = 1, 2, \cdots, m$, the $\mathbf{v}$ is the vector of the views, the m is the number of the views.

In our model, the m is 4 because we have four representations of rating network structure and review texts for users and items totally. In our experiments, we use a simple expression shown in [1] that is equivalent to Multi-view Machine as discussed above.

3 Experiments

In this section, we do extensive experiments to validate the effectiveness of the proposed RLDB on three real datasets, compared to the state-of-the-arts.

3.1 Dataset

The experiments are conducted on three representative datasets.

- Amazon[4]: Amazon review datasets [6] contains product reviews and metadata from Amazon website. In our experiment, we select three of the largest categories including Instant Video, Electronics and Home, called Amazon_V, Amazon_E, and Amazon_H.
- Yelp[5]: It is a large-scale datasets consisting of restaurant reviews.
- TripAdvisor[6]: TripAdvisor is an American travel website company providing hotels booking as well as reviews of travel-related content.

In our experiments, we adopt the widely used criterion of Mean Squared Error (MSE) [6] to evaluate the performance of our proposed model.

$$MSE = \frac{1}{N} \sum_{n=1}^{N} (r_n - \hat{r}_n)^2, \tag{8}$$

where r_n is the nth gold rating score, rn is the nth predicted rating score and N is the total number of ratings.

3.2 Compared Methods

Compared methods involved in experiments are summarized as follows.

- MF [4]: Matrix Factorization is the important popular CF-based recommendation method. It only uses rating matrix as input and estimates two low-rank matrices to predict ratings.
- HFT [6]: Hidden Factor as Topic proposed in employs topic distributions to learn latent factors from user or item reviews.
- ConvMF [3]: ConvMF integrates CNN into probabilistic matrix factorization which captures contextual information of documents.
- DeepCoNN [9]: DeepCoNN adopts two parallel CNN to model user behaviors and item properties from all the review texts.

3.3 Effectiveness Experiments

The performances of RLDB and the compared methods are reported in Table 1 with the best performance shown in bold. From the Table 1, we have the following observations.

The proposed model RLDB outperforms other approaches on all the datasets. Since the CF-based method MF only use the rating score information, it has poor performance on the rating prediction. Although DeepCoNN [9] adopt deep

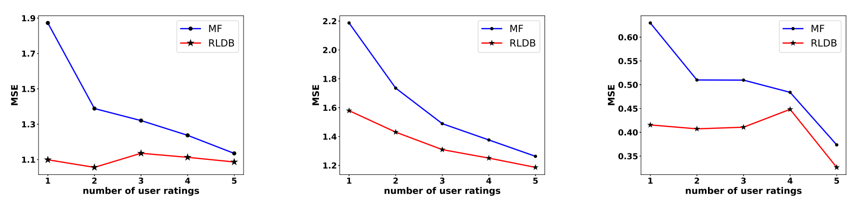
[4] `https://snap.stanford.edu/data/web-Amazon.html`
[5] `https://www.yelp.com/dataset-challenge`
[6] `https://www.tripadvisor.com/`

Table 1: **MSE Comparison with other methods**. **Best results are indicated in bold**.

Dataset	MF	HFT-10	ConvMF	DeepCoNN	RLDB
Amazon_V	1.5319	1.5470	1.4159	1.1349	**1.0755**
Amazon_E	1.7918	1.6137	1.9550	1.6331	**1.5139**
Amazon_H	1.4737	1.5103	1.5040	1.4199	**1.3202**
Yelp	1.6129	1.6064	1.9570	1.3982	**1.3160**
TripAdvisor	0.5722	0.4071	0.4895	0.4113	**0.3962**

learning model, its performance is still worse than our RLDB due to the use of single-view data (review information). The ConvMF and RLDB both utilize two views of data. The better performance of RLDB shows the more effective combination of multi-view data in RLDB. In all, the good performance of RLDB attribute to the effective feature extraction with representation learning from multi-view data and fusion mechanism with multi-view machine in RLDB.

3.4 Cold Start Experiment



(a) Amazon_V (b) Yelp (c) TripAdvisor

Fig. 3: MSE of RLDB compared to MF for users with different number of training reviews

In this section, we evaluate the MSE score on five types of cold start users with different numbers of rated items (e. g. , users with the number of rated items no more than 5). As a representative of CF, MF is included to compare their performance. The results are shown in Fig. 3. We can observe that the RLDB always performs better than MF. More importantly, the superiority of RLDB is more significant for more cold users (i. e. , users with less rating interactions). It shows that RLDB has the potential to alleviate the cold start problem. We think it attributes to multi-view data and the delicate design of RLDB: representation learning model with depth and breadth.

4 Conclusion

In order to make full use of rich multi-view data in recommender system, we propose a novel model RLDB based on representation learning. The RLDB designs a heterogeneous information network embedding method and CNN to learn representation of users and items from structure and text information, respectively. Furthermore, a multi-view machine is employed to effectively fuse these features. Extensive experiments on real-world datasets show RLDB can boost the recommendation performance and benefit for cold start problem.

5 Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61772082, 61375058), the National Key Research and Development Program of China (2017YFB0803304), and the Beijing Municipal Natural Science Foundation (4182043).

References

1. Cao, B., Zheng, L., Zhang, C., Yu, P.S., Piscitello, A., Zulueta, J., Ajilore, O., Ryan, K., Leow, A.D.: DeepMood: Modeling mobile phone typing dynamics for mood detection. In: SIGKDD (2017)
2. Cao, B., Zhou, H., Li, G., Yu, P.S.: Multi-view machines. In: WSDM. pp. 427–436. ACM (2016)
3. Kim, D., Park, C., Oh, J., Lee, S., Yu, H.: Convolutional matrix factorization for document context-aware recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems. pp. 233–240. ACM (2016)
4. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer 42(8) (2009)
5. 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. ACM (2014)
6. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on Recommender systems. pp. 165–172. ACM (2013)
7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: NIPS. pp. 3111–3119 (2013)
8. Shi, C., Li, Y., Zhang, J., Sun, Y., Philip, S.Y.: A survey of heterogeneous information network analysis. IEEE Transactions on Knowledge and Data Engineering 29(1), 17–37 (2017)
9. Zheng, L., Noroozi, V., Yu, P.S.: Joint deep modeling of users and items using reviews for recommendation. In: WSDM. pp. 425–434. ACM (2017)