Align Reviews with Topics in Attention Network for Rating Prediction

Yile Liang [1], Tieyun Qian✉ [1,2], Huilin Yu [1]

[1] School of Computer Science, Wuhan University, Hubei, China
[2] Corresponding author
{liangyile, qty, huilin_yu}@whu.edu.cn

Abstract. Rating prediction has long been a hot research topic in recommendation systems. Latent factor models, in particular, matrix factorization (MF), are the most prevalent techniques for rating prediction. However, MF based methods suffer from the problem of data sparsity and lack of explanation. In this paper, we present a novel model to address these problems by integrating ratings and topic-level review information into a deep neural framework. Our model can capture the varying attentions that a review contributes to a user/item at the topic level. We conduct extensive experiments on three datasets from Amazon. Results demonstrate our proposed method consistently outperforms the state-of-the-art recommendation approaches.

Keywords: rating prediction; topic model; deep learning

1 Introduction

Rating prediction aims to predict the user's ratings on unrated items which may reflect the user's potential interests towards the item. Latent factor models, especially matrix factorization (MF), which are widely used techniques towards this problem, suffer from the severe data sparsity problem when the number of items in the platform becomes extremely large.

In order to address the sparsity issues, researchers have devoted extensive efforts to leverage various types of side information. Among which, the users' reviews attract lots of research interests. A recent trend is applying deep learning techniques to rating prediction [1, 2, 4–6]. These approaches differ mainly in how they utilize textual information.

In this paper, we present a novel deep neural framework which is *topic-oriented* and *can selectively focus on informative reviews*. To this end, we first apply the simplest LDA topic model to each review to get the review's initial topic distribution. We then design an attention network which can assign reviews' topical importance to users/items by automatically learning the reviews' attention weights. We finally present a neural prediction layer to include both the latent factors from ratings and textual information from reviews.

2 Our Proposed Model

The architecture of proposed ARTAN (**A**lign **R**eviews with **T**opics in **A**ttention **N**etwork) model is shown in Fig. 1.

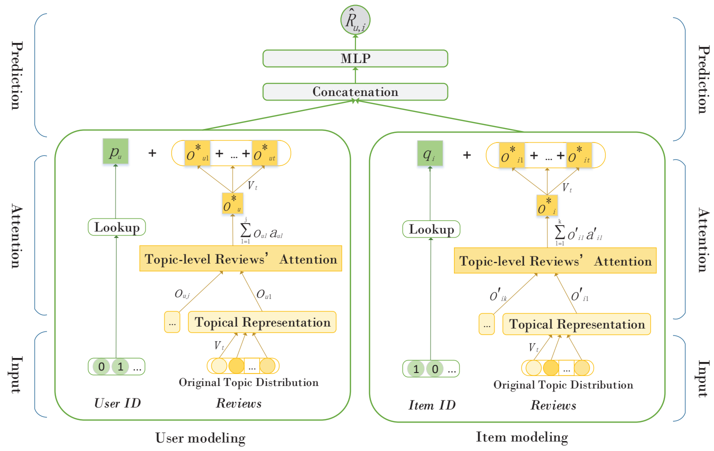

Fig. 1. The architecture of ARTAN

Input Layer. *Initial Latent Factor Representation of User and Item.* We set up a lookup table to transform the one-hot representations of userID and itemID into low-dimensional dense vectors as $\mathbf{p_u} \in \mathbb{R}^K$, $\mathbf{q_i} \in \mathbb{R}^K$.

Topical Representation of Review. We input each review j into the LDA model to obtain its topical distribution θ_j with two hyper-parameters $\alpha = 0.25$ and $\beta = 0.1$. We introduce a matrix $\mathbf{V} \in \mathbb{R}^{T\times K}$ to associate the embeddings to different topics, where T is the number of topics and K is the dimensionality. The topical representation of review j of user u can be calculated by aggregating the weighted embeddings over topics:

$$\mathbf{o_{u,j}} = \sum_{t=1}^{T} \theta_{j,t} \cdot \mathbf{v_t} \tag{1}$$

Attention Layer. The input of the attention network consists of the topical representation of user u's j-th review ($\mathbf{o_{u,j}}$) and a randomly initialized vector ($\mathbf{a'_u}$). The user u's attention weight on a review j is defined as:

$$a^*_{u,j} = \mathbf{h^T} \cdot ReLU(\mathbf{W_o o_{u,j}} + \mathbf{W_u a'_u} + \mathbf{b}), \tag{2}$$

where $\mathbf{h} \in \mathbb{R}^K$, $\mathbf{W_o} \in \mathbb{R}^{K\times K}$, $\mathbf{W_u} \in \mathbb{R}^{K\times K}$ and $\mathbf{b} \in \mathbb{R}^K$, are parameters that project the input into the hidden layer. The final attention $a_{u,j}$ is obtained by applying the softmax function to the original $a^*_{u,j}$ as normalization. After that, the feature vector of user u is calculated as the weighted sum of all reviews:

$$a_{u,j} = \frac{exp(a^*_{u,j})}{\sum_{j=1}^{l} exp(a^*_{u,j})}, \qquad \mathbf{o^*_u} = \sum_{j=1}^{l} a_{u,j} \cdot \mathbf{o_{u,j}} \tag{3}$$

To further strengthen the impacts of topics, we let the above user feature vector $\mathbf{o^*_u}$ interact with each topic embedding $\mathbf{v_t}$. In this way, the topical representation of user u can be highlighted and the user's interests would be associated with each topic once again.

$$\mathbf{o_u} = \sum_{t=1}^{T} \mathbf{o^*_u} \odot \mathbf{v_t}, \tag{4}$$

where $\odot$ denotes the element-wise product.

Prediction Layer. Given a user u, an item i and their reviews, we now take the latent factor representations on ratings and topical representations on reviews to model the interaction between user u and item i as follows.

$$\mathbf{x}^* = (\mathbf{p_u} + \mathbf{o_u}) \oplus (\mathbf{q_i} + \mathbf{o_i}), \tag{5}$$

where $\oplus$ denotes the concatenation operation. We present a multi-layer perceptron (MLP) based structure on $\mathbf{x}^*$ for modeling the complicated interactions and getting the real-valued rating $\hat{R}_{u,i}$.

Network Training. Since our target is rating prediction, we treat it as a regression task and adopt the square loss as the objective function:

$$\mathcal{L} = \sum_{u,i \in \mathcal{T}} (\hat{R}_{u,i} - R_{u,i})^2, \tag{6}$$

where $\mathcal{T}$ denotes the set of instances for training, i.e.,$\mathcal{T} = \{(u,\ i,\ r_{u,i},\ d_{u,i})\}$, $R_{u,i}$ is ground truth rating in training set $\mathcal{T}$, and $\hat{R}_{u,i}$ is the predicted rating.

3 Experiments

Datasets and Experimental Settings. We use three publicly available datasets from Amazon[3] to evaluate our model, including Patio Lawn and Garden, Automotive, and Grocery and Gourmet Food. We take the 5-core version for experiments [2, 1, 6]. We randomly split the datasets into training/validation/test sets with a 80/10/10 split and use MSE as the evaluation metric. The statistics of these datasets are shown in Table 1.

Table 1. Statistics of the evaluation datasets

Datasets	#users	#items	#ratings	#sparsity
Garden	1686	962	13272	0.9918
Automotive	2928	1834	20473	0.9962
Grocery	14679	8711	151254	0.9988

We choose six state-of-the-art methods as our baselines, including *SentiRec[4]*, *TARMF[5]*, *MPCN[6]*, *NARRE[1]*, A^3*NCF[2]* and *ALFM[3]*. For the baselines, we use the same hyper-parameter settings if they are reported by authors. For our ARTAN model, we set the number of topics $T = 25$, the number of latent factors K is 25, and the learning rate is 0.001.

Performance Evaluation. The performance of ARTAN and the baselines on all datasets are reported in Table 2. From the results, we have the following important observations.

Firstly, our ARTAN model significantly outperforms the state-of-the-art baselines on all the datasets.

Secondly, among four word-based methods (SentiRec, TARMF, MPCN, NARRE), NARRE is the best in consideration of the different weights of reviews. While the word-level textual features are discentralized, its performance is inferior to our model. Not utilizing users' and items' textual representations at testing stage in

[3] http://jmcauley.ucsd.edu/data/amazon/links.html

Table 2. Performance comparison on four datasets. The best performance among all is in bold while the best one among baselines is marked with an underline. * denotes the index of the baselines.

Method	Index	Results(MSE)			Improvements(%) – (7) vs. *		
		Garden	Automotive	Grocery	Garden	Automotive	Grocery
SentiRec	(1)	1.0671	0.8245	1.0140	9.79	7.50	2.56
TARMF	(2)	1.1025	0.8686	1.0733	12.69	12.19	7.95
MPCN	(3)	1.1664	0.8154	1.0941	17.47	6.46	9.70
NARRE	(4)	0.9904	0.7815	<u>0.9972</u>	2.81	2.41	0.92
A^3NCF	(5)	1.0349	0.8228	1.0199	6.99	7.30	3.13
ALFM	(6)	<u>0.9835</u>	<u>0.7718</u>	1.0018	2.13	1.18	1.38
ARTAN	(7)	**0.9626**	**0.7627**	**0.9880**	–	–	–

TARMF, and only selecting the most representative reviews in MPCN lead to the poor result. The poor performance of SentiRec can be due to it not differentiating reviews at all.

Thirdly, two topic-based methods (A^3NCF, ALFM) are not as good as our model since they neglect the difference among various reviews when building topic model. In contrast, our ARTAN is not only topic oriented, but also assigns varying weights for different reviews with attention mechanism, and thus achieves the best performance.

4 Conclusion

In this paper, we propose a novel model for rating prediction. Our model adopts the deep neural attention network as architecture to extract representations from reviews in addition to latent factors from ratings. Different from previous topic level studies in utilizing reviews, our model can automatically identify the important reviews for each user/item with the attention mechanism. We conduct extensive experiments on three real world datasets. Results demonstrate that our model significantly outperforms the state-of-the-art baselines.

Acknowledgment

The work described in this paper has been supported in part by the NSFC project (61572376).

References

1. Chen, C., Zhang, M., Liu, Y., Ma, S.: Neural attentional rating regression with review-level explanations. In: WWW. pp. 1583–1592 (2018)
2. Cheng, Z., Ding, Y., He, X., Zhu, L., Song, X., Kankanhalli, M.: A^3ncf: An adaptive aspect attention model for rating prediction. In: IJCAI. pp. 3748–3754 (2018)
3. Cheng, Z., Ding, Y., Zhu, L., Kankanhalli, M.: Aspect-aware latent factor model: Rating prediction with ratings and reviews. In: WWW. pp. 639–648 (2018)
4. Hyun, D., Park, C., Yang, M.C., Song, I., Lee, J.T., Yu, H.: Review sentiment-guided scalable deep recommender system. In: SIGIR. pp. 965–968 (2018)
5. Lu, Y., Dong, R., Smyth, B.: Coevolutionary recommendation model: Mutual learning between ratings and reviews. In: WWW. pp. 773–782 (2018)
6. Tay, Y., Luu, A.T., Hui, S.C.: Multi-pointer co-attention networks for recommendation. In: KDD. pp. 2309–2318 (2018)