A Deep Recommendation Model Incorporating Adaptive Knowledge-Based Representations

Chenlu Shen[1], Deqing Yang[1(✉)], and Yanghua Xiao[2,3]

[1] School of Data Science, Fudan University, Shanghai 200433, China
{clshen17,yangdeqing}@fudan.edu.cn
[2] School of Computer Science, Fudan University, Shanghai 200433, China
shawyh@fudan.edu.cn
[3] Shanghai Institute of Intelligent Electronics and Systems, Fudan University, Shanghai 200433, China

Abstract. Deep neural networks (DNNs) have been widely imported into collaborative filtering (CF) based recommender systems and yielded remarkable superiority, but most models perform weakly in the scenario of sparse user-item interactions. To address this problem, we propose a deep knowledge-based recommendation model in which item knowledge distilled from open knowledge graphs and user information are both incorporated to extract sufficient features. Moreover, our model compresses features by a convolutional neural network and adopts memory-enhanced attention mechanism to generate adaptive user representations based on latest interacted items rather than all historical records. Our extensive experiments conducted against a real-world dataset demonstrate our model's remarkable superiority over some state-of-the-art deep models.

1 Introduction

Despite of the superior performance of DNNs, some challenges are still not tackled well by previous deep recommendation models. The first one is *sparse observed user-item interactions*. In general, CF-based methods such as [2,4] suffer from few observed interactions, also known as the problem of *cold start* or *data sparsity*. The second one is *dynamic and diverse preferences of users*. In many deep recommendation models [2,3], user representations are generated based on all historical interacted items. However, user interests may shift as time elapses hence the latest records are more important than the early ones for preference inference. On the other hand, a user's representation is usually fixed which is not adaptive to the diversity of his/her preferences. To address above issues, we propose a novel deep recommendation model which not only imports external knowledge but also integrates some deep components as follows.

This work was supported by Chinese NSFC Project (No. U1636207, No. 61732004).

G. Li et al. (Eds.): DASFAA 2019, LNCS 11448, pp. 481–486, 2019.
https://doi.org/10.1007/978-3-030-18590-9_71

First, item representations are enriched by the knowledge distilled from open knowledge graphs (KGs) resulting in immunity from sparse observed user-item interactions, and high interpretability of recommendation results. In addition, user's *personal attribute embedding* is generated based on personal information, e.g., demographics or tags.

Second, our model also generates user's *historical preference embedding* to constitute comprehensive user representation, which is a temporal and adaptive vector generated through fusing latest interacted item embeddings by *memory-enhanced attention mechanism*.

Third, for predicting the probability that a given user likes a candidate item which is the key of achieving top-N recommendation, our model further utilizes a multi-layer perceptron (MLP) fed with element-wise product of the user embedding vector and the item embedding vector to generate a non-linearity output better than traditional inner products [2].

In summary, our contributions in this paper include:

1. We propose a deep knowledge-based recommendation model to overcome the sparsity of user-item interactions with external knowledge from KGs as auxiliary information.
2. We build a memory-enhanced attention mechanism which is helpful to generate user representations adaptive to their dynamic and diverse preferences.
3. Extensive experiments are conducted on Douban[1] movie dataset to justify that our model outperforms some state-of-the-art models.

2 Model Description

Without loss of generality, we explain the details of our model w.r.t. Douban movie recommendation in this section. Figure 1 depicts the framework of our model.

Knowledge-Aware Item/User Embedding. We first leverage knowledge as item features to discover latent relationships between items. A movie i's representation is projected into a knowledge-based tensor $\boldsymbol{P}_i \in \mathbb{R}^{f \times e}$, where f is feature number and e is embedding dimension. That is, each row of $\boldsymbol{P}_i$ relates to a feature embedding. Then, since each Douban user is specified by a set of tags, a user u's *personal attribute embedding* $\boldsymbol{p}_u^a$ is the average of u's tag embeddings. At the same time, we employ attention mechanism to generate u's *historical preference embedding* $\boldsymbol{p}_{uj}^h$ which is adaptive to different candidate movie j. Then, we get u's comprehensive representation w.r.t item j as follows,

$$\boldsymbol{p}_{uj} = \gamma \boldsymbol{p}_{uj}^h + (1-\gamma)\boldsymbol{p}_u^a \tag{1}$$

where γ is a controlling parameter.

[1] https://movie.douban.com, a famous Chinese website of movie reviews.

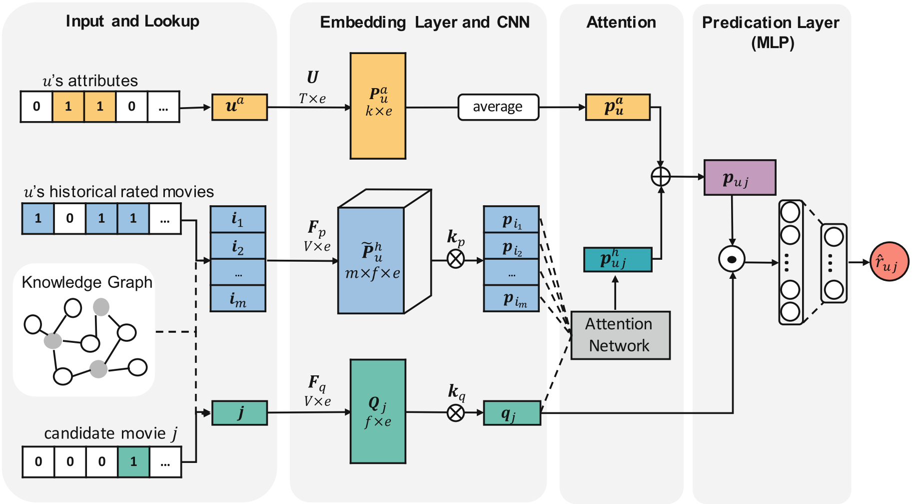

Fig. 1. The framework of our deep knowledge-based recommendation model.

Convolutional Neural Network for Embedding Compression. Given the shifted user preference, we only pay attention to the m latest interacted items to represent a user's preference. The union of m rated movies' representations is a feature cube (tensor) $\tilde{\boldsymbol{P}}_u^h \in \mathbb{R}^{m\times f\times e}$. Then, we use CNN for compressing features and finally get a squeezed matrix $\boldsymbol{P}_u^h = [\boldsymbol{p}_{i_1}, \boldsymbol{p}_{i_2}, ..., \boldsymbol{p}_{i_m}] \in \mathbb{R}^{m\times e}$. Specifically, the convolutional operation is the inner product without activation.

Memory-Enhanced Attention Mechanism. Suppose $\boldsymbol{p}_i$ and $\boldsymbol{q}_j$ are the embedding of historical movie i and candidate j respectively, we adopt a two-layer fully connected neural network to calculate their similarity

$$sim(\boldsymbol{p}_i, \boldsymbol{q}_j) = \boldsymbol{g}^\top ReLU\big(\boldsymbol{W}(\boldsymbol{p}_i \odot \boldsymbol{q}_j) + \boldsymbol{b}\big), i \in \mathcal{M}_u^+ \tag{2}$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ are the weight matrix and bias respectively, and $\boldsymbol{g}$ is a weight vector for output projection, $\odot$ denotes the element-wise product operator. As we emphasized before, we capture the historical preference of a user through a *memory component* that stores a fixed number of latest interacted items at time t as $\mathcal{M}_u^+ = \{i_1^t, i_2^t, ..., i_m^t\}$. Then, the attention weight α_{ij} and u's historical preference embedding are computed as follows

$$\alpha_{ij} = \frac{\exp\big(sim(\boldsymbol{p}_i, \boldsymbol{q}_j)\big)}{\sum_{i\in\mathcal{M}_u^+} \exp\big(sim(\boldsymbol{p}_i, \boldsymbol{q}_j)\big)}, \qquad \boldsymbol{p}_{uj}^h = \sum_{i\in\mathcal{M}_u^+} \alpha_{ij}\boldsymbol{p}_i \tag{3}$$

Prediction Layer. For final prediction, we adopt a MLP fed with user embeddings and item embeddings to enhance the capability of capturing non-linearity.

Each hidden layer is activated by *ReLU* and output layer is activated by sigmoid function for final prediciton $\hat{r}_{uj}$.

$$\boldsymbol{y}_l = ReLU(\boldsymbol{W}_l \boldsymbol{y}_{l-1} + \boldsymbol{b}_l), \qquad \hat{r}_{uj} = \sigma(\boldsymbol{h}^\top \boldsymbol{y}_L + b) = \frac{1}{1 + e^{\boldsymbol{h}^\top \boldsymbol{y}_L + b}} \tag{4}$$

Model Learning. Our model belongs to binary classifier based on implicit feedback [3]. Each sample is formalized as a triplet $<u, j, r_{uj}>$. We use Adagrad [1] to optimize the objective function, which is the binary cross-entropy with regularization as a classic objective for training a neural network classifier.

3 Experiment

Dataset Description. We crawled Douban movie dataset from its website, including ratings and user tags, which has been released[2]. We also distilled movie knowledge from an open Chinese KG named *CN-DBpedia* [7].

Sample Collection and Evaluation Metrics. We conducted our experiments on different sparsity levels, according to a certain proportion s by sampling. Specifically, we evaluated model performance as s increases gradually.

Compared Models. *BPR* [5] is a MF-based model uses Bayesian Personalized Ranking loss as a ranking-ware objective function. A widely used deep model *NCF* [3] updates randomly initialized latent embeddings by observed interactions. *FISM* [4] is a representative factored item-based CF model that learns item similarity matrix as the product of two low dimensional latent factor matrices. *NAIS* [2] tailors attention mechanism to differentiate the weights of historical items in user representation extended from FISM. Then, our model is denoted as *KAC* (Knowledge+Attention+CNN) in the following texts. We further propose four variants to be compared which are named as *AC/KC/KAavg/KAcon*. Particularly, KAavg and KAcon replace CNN with average and concatenate.

Experiment Results. With the optimal settings of hyper parameters, we compared KAC with other models. Figure 2 displays top-1/3/5 movie recommendation performance of three evaluation metrics, i.e., HR (Hit Ratio), nDCG (Normalized Discounted Cumulative Gain), AP (Average Precision) and RR (Reciprocal Rank). The results show that KAC outperforms others in all scenarios, and the variants also prevail against the baselines. KAC's superiority justifies the significance of adaptive knowledge-based embeddings. Specifically, KAC's superiority over KA/AC/KAavg/KAcon shows the effects and necessities of corresponding components in our model.

---

[2] http://gdm.fudan.edu.cn/GDMWiki/Wiki.jsp?page=Network%20DataSet.

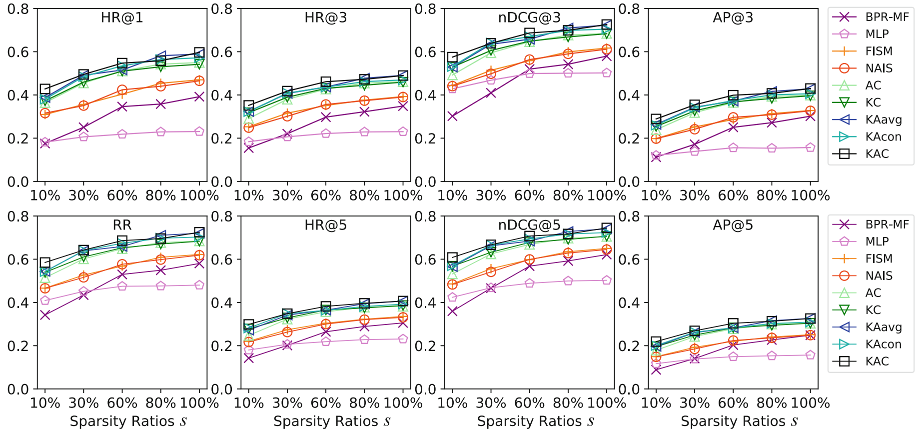

Fig. 2. Performance comparisons of recommendation with different sparsity ratios.

4 Related Work and Conclusion

As a pioneer work of deep recommendation model, NCF [3] fuses a generalized MF layer and an MLP together to learn user/item embeddings based on observed interactions. DKN [6] incorporates knowledge information of entities in news contexts from KGs. On the other hand, NAIS [2] pays adaptive attention to historical items for user representation according to different candidate items, which has been proven effective by empirical studies.

Comparatively, we propose a knowledge-based deep model in this paper, which is designed towards the recommendation of sparse user-item interactions. Our model not only strengthens user/item representation by importing knowledge and users' personal information to alleviate the sparsity problem, but also model employs an enhance attention mechanism with a memory component to better capture dynamic user preferences. We also adopt CNN to extract and compress user features in our model. The results of extensive evaluation not only justify our model's superiority but also confirm the significance of incorporating related DNN-based components.

References

1. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. **12**(7), 257–269 (2011)
2. He, X., He, Z., Song, J., Liu, Z., Jiang, Y.G., Chua, T.S.: Nais: neural attentive item similarity model for recommendation. IEEE TKDE **30**, 2354 (2018)
3. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: Neural collaborative filtering. In: Proceedings of WWW (2017)
4. Kabbur, S., Ning, X., Karypis, G.: Fism: Factored item similarity models for top-n recommender systems. In: Proceedings of KDD (2013)

5. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of CUAI (2009)
6. Wang, H., Zhang, F., Xie, X., Guo, M.: DKN: deep knowledge-aware network for news recommendation (2018)
7. Xu, B., et al.: CN-DBpedia: a never-ending Chinese knowledge extraction system. In: Benferhat, S., Tabia, K., Ali, M. (eds.) IEA/AIE 2017. LNCS (LNAI), vol. 10351, pp. 428–438. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60045-1_44