

Hybrid Recommendation Algorithms Based on ConvMF Deep Learning Model

Jiakun Zhao, Zhen Liu*, Huimin Chen, Jingbo Zhang and Qing Wen

School of software, Xi'an Jiao-tong University, Shaanxi, 710049, China

*Corresponding author

Abstract—Due to ConvMF (Convolutional Matrix Factorization) use side information to improve the accurate of the prediction rating, it shows side information is important for rating prediction accuracy. but it does not make fully use of the features of the item description documents such as reviews, abstract, or synopses. To handle the problem, this paper proposes a novel model DE-ConvMF, which have double embedding layer in ConvMF, take more attention on the item side information. This double embeddings includes two part: one is general embedding layer, other is domain embedding layer, we combine general embedding with domain embedding as the embedding layers. Then we use Stack Denoising Auto Encoder (SDAE) to deal with users side information(age, sex, occupation), Through the user ratings and labels to improve the accuracy of forecast scores. Extensive experiment results on movielens (ml-10M) datasets show that our new model outperforms other methods in effectively utilizing side information and achieves performance improvement.

Keywords—recommender systems; ConvMF; DE-CNN; SDAE

I. INTRODUCTION

Sparseness of user-to-item rating data is one of the major factors that deteriorate the quality of recommender system. Recently, some approaches use side information to generating more accurate model which improve the rating prediction accuracy. Specifically, Kim et al. proposed Convolutional Matrix Factorization (ConvMF)[1] that integrates Convolutional Neural Networks (CNN)[2] into probabilistic matrix factorization(PMF)[3]. For MovieLens datasets, the improvements of ConvMF over the best competitor, collaborative deep learning (CDL)[4], are 3.92% on ML-1m dataset and 2.79% on ML-10m dataset. It is undoubtedly the best predictive scoring model at present. ConvMF successfully captures subtle contextual difference Information of documents such as surrounding words and words orders. For example, suppose that the following two sentences are given in a document: “what is your aim in life”, “I aim to be a successful writer.” Although each occurrence of “aim” seems to have almost the same meaning, there is a subtle syntactic difference between these words--a noun and a verb. ConvMF easily to distinguish those subtle change. Because of the CNN quickly captures local features of documents through modeling components such as local receptive fields, shared weights.

Existing deep learning models for aspect extraction use either a pre-trained general-purpose embedding, such as GloVe[5], or a general review embedding. As we can see on the

paper of ConvMF, with pre-trained word embedding model like ConvMF+, and general review embedding model like ConvMF. However, aspect extraction is a complex task that also requires fine-grained domain embeddings. For example, from “Its speed is amazing” in a laptop review, it aims to extract “speed”. if we not in certain domain, it will mean different. “speed” has a very special meaning (how many instructions per second) in the laptop domain, whereas “speed” in general embedding or general review embedding may mean how many miles per second. So using domain embedding is important even when the domain embedding corpus is not large. Thus, Hu Xu proposed a new CNN model named Dual Embeddings CNN (DE-CNN)[6]. To address aspect extraction on recommender system, we proposed a new model based on ConvMF, called DE-ConvMF which Integrate the advantage of the DE-CNN into ConvMF.

However, the new model does not utilize user side information. We use deep learning technology SDAE to process user tag information. Take some noise to auxiliary information, as the input of the model, through encoder and decoder, get the L/2 layers weights and biases, as user latent model input.

This paper proposes two different models, one is DE-ConvMF, which use double embedding to enhance the ratings prediction. The other is SDAE-DE-ConvMF, which use users side informations to get more better recommend. The rest of the paper is organized as follows. Section 2 briefly reviews the mainly relevant technologies and approaches. Section 3 introduces our new model DE-ConvMF and SDAE-DE-ConvMF. Section 4 experimentally evaluates. Finally, give the conclusion and future work in Section 5.

II. RELEVANT TECHNOLOGIES AND APPROACHS

In this section, we will briefly introduce Convolutional Matrix Factorization(ConvMF), double embedding Convolutional Neural Network(DE-CNN), Stack Denoising Auto Encoder (SDAE).

ConvMF: ConvMF is a popular context-aware model in recommendation systems.it integrates convolutional neural network(CNN) into probabilistic matrix factorization(PMF). So that ConvMF captures contextual information of documents and further enhances the rating prediction accuracy. This model use the movie plot or summary as the input of the CNN, via the embedding layer that make the movie plot or summary convert to word vector. Then, fixed the dimension of the word vector, let the embedding layer output as the input of the convolution layer. Next, the pooling layer extracts representative features from the

convolution layer and deals with variable lengths of documents via pooling operation that constructs a fixed-length feature vector. Finally, generally speaking, at the output layer, high-level features obtained from the previous layer should be converted for a specific task. Thus, through projection operation, produces a document fixed-dimensional latent vector by using conventional nonlinear projection. Eventually, CNN architecture becomes a function that input a movie plot or summary, and output a latent vectors of each movies. Meanwhile, set the latent user models and movie latent models project on a k-dimensional space for recommendation task.

DE-CNN: DE-CNN is a new model on CNN for aspect extraction. The DE-CNN model consists mainly of two embedded layers, four convolutional neural network layers, and a fully connected layer shares the output of the convolutional layer. Existing deep learning models for aspect extraction use a pre-trained embedding like GloVe or a general review embedding. It address the automated feature learning and how to use a more simple model to deal with a complex problem. This model proposed a double embedding mechanism that is shown crucial for aspect extraction. One embedding is about general embedding, the other is domain embedding. It leverage both general embedding and domain embedding and let the rest of the CNN architecture to decide which embedding have more useful information.

SDAE: SDAE is a feedforward deep neural network that requires multiple noise reduction self-encoding processes, with each layer of output as input, where the input of the first layer is the corrupted original data. After several times of hidden layer encoding and decoding process, the final output is obtained. Each layer in the stack noise reduction encoder is independently unsupervised, and the input of each layer is the output of the previous layer, except that the input of the first layer is the raw data after the noise is added. Layer-by-layer greedy training, with minimal input and minimum error after reconstruction. After training the front L layer, train the L+1 layer again. Because it is a feedforward neural network, use the output of the L-th layer. The input of the L+1th layer is used to train the L+1th layer. In the training, the parameter model of the former L layer has reached the optimal solution. After the layer-by-layer training is over, a fine-tuning process is also required. SDAE is a popular deep learning model. Its advantage is that it has the characteristics of fitting arbitrary complex functions. There are many parameters in this model, so the hypothesis spatial dimension to be obtained is very high, and has a strong representation ability, and the multi-layer network extraction features of the model are more vivid and representative.

III. DE-CONVMF MODEL AND SDAE-CONVMF MODEL

In this part, we will respectively introduce the details about our two new models. First of all, we introduce the first new model of DE-ConvMF. Then, based on DE-ConvMF, we proposed a new model named SDAE-DE-ConvMF.

DE-ConvMF: DE-ConvMF is the new model we proposed. It combine DE-CNN with ConvMF, to enhance get the representation feature of the movie plot. Although the ConvMF

model also expresses the relationship between words and words, it does not take into account the difference in meaning between words in the general field and specific fields. The general field can only roughly represent the word vector of a word, and the word vector of a specific field can more express the word vector specific to the word in a certain field. For example, we say "it's incredibly fast", "it" is a generic word, and general embedding may be a good representation of the word vector. However, "speed" has a very fine-grained meaning in our particular computer field, that is, how many instructions can be executed per second, and for the general field of embedded it may mean how many kilometers per second can run. Therefore, we use general domain embedding and domain-specific embedding to let the rest of the network decide which embedding have more important information. Based on this problem, we bring the two-stage embedding layer into the ConvMF model, and form a new improved model DE-ConvMF, as shown in Figure 1:

In the Fig 1, this paper first preprocesses the embedded layer data, which is mainly divided into two parts: 1. General Embedding indicates that the Stanford-designed GloVe Wikipedia word vector is used. 2. Domain Embedding indicates that the GloVe algorithm model is used to train the word vector of the movie introduction text, which is more specific to the film field than the Wikipedia word vector. This article is to accumulate the dimensions of both. For example, the length of the profile text information of the movie M is L, the word vector of the general field of the word i is W_i^G , and the word vector of the specific domain is W_i^D , and the profile text information of the movie M can be expressed in the form of a matrix $A \in \mathbb{R}^{p \times L}$, where $p=p_1+p_2$, p_1 is the dimension of the word vector of the general domain, and p_2 is the dimension of the word vector of the specific domain. In the pooling layer, we use avgpooling replace maxpooling. Then, the rest of parameters and architecture of DE-ConvMF like ConvMF, do not change.

SDAE-DE-ConvMF: Since the DE-ConvMF model only uses the auxiliary information of the movie introduction text to improve the accuracy of the score prediction, there is no potential auxiliary information related to the user's label in the user's potential model. We will use the SDAE model to explore the potential auxiliary information of users and use it to improve the accuracy of predicting movie scores. Because the DE-ConvMF model of this paper only conducts in-depth research on the text information of the movie introduction, and uses it to improve the accuracy of the prediction score, the auxiliary label information such as the user's age, occupation, gender, etc. is not utilized. The new model SDAE-DE-ConvMF uses SDAE technology to extract the characteristics of the client, and combines the DE-ConvMF model of this paper to improve the accuracy of the prediction score. Through the introduction in the previous section, we learned that SDAE technology is aimed at feature extraction, and can be seamlessly connected with the probability matrix decomposition model. Next, this article will introduce the new model SDAE-DE-ConvMF in detail. In order to apply the auxiliary information of the user tag to the new model, the model of SDAE-DE-ConvMF designed in this paper is shown in Figure 2.

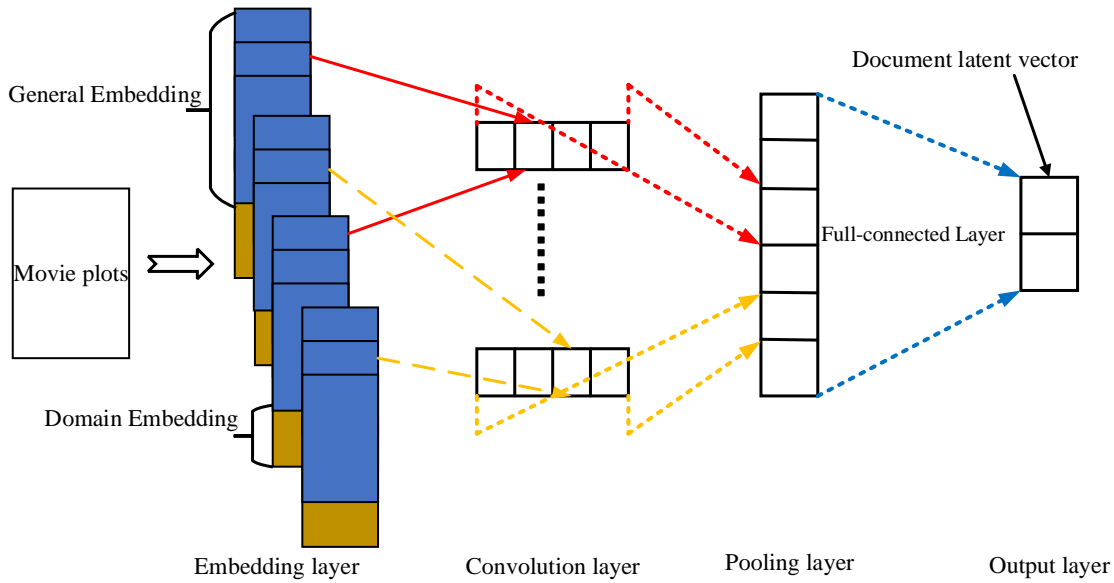


FIGURE I. OUR DE-CNN ARCHITECTURE FOR DE-CONVMF

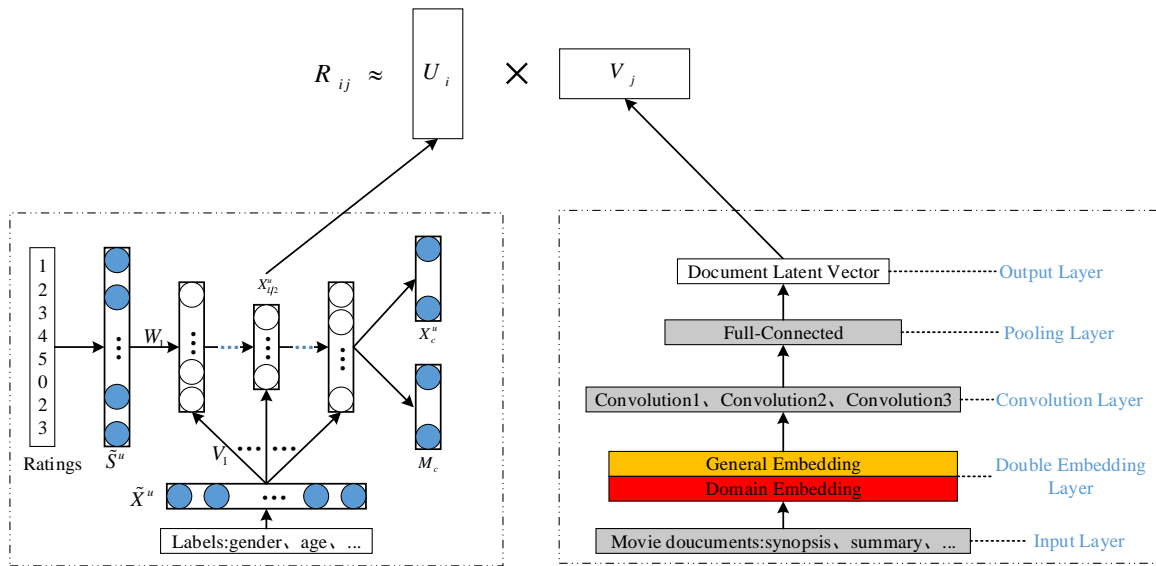


FIGURE II. SDAE-DE-CONVMF MODEL ARCHITECTURE

The left side of Fig. 2 represents the potential model part of the user, using the SDAE model to extract features, the right part representing the potential model of the movie, and the DE-ConvMF model to extract the features of the movie text.

For the potential model of the user, this article takes the user's rating and the user's tag information as input to the SDAE, where the user's rating $R_{ij} \in \mathbb{R}^{m \times n}$. In this paper, the score is divided into movies, and a data set $S^u \in \mathbb{R}^{m \times n}$ can be obtained. This data set contains m users, which can be expressed as s , and each user i can score n movies, which can be expressed as

$S_i^u = \{R_{i1}^u, R_{i2}^u, R_{i3}^u, \dots, R_{in}^u\}$. The potential model optimizes the objective function of the following formula by obtaining the R_{ij} , the score information matrix S^u and the potential factor of the user tag auxiliary information $X_i \in \{X_1, X_2, X_3, \dots, X_n\}$:

$$\min_U L_R(R, UV^T) + \lambda \|U\|_F^2 + \alpha f(S^u, U, X) \quad (1)$$

The new model SDAE-DE-ConvMF, the hidden layer serves as a bridge between user ratings and user tag assistance information, with the front layer acting as an encoder and the back layer acting as a decoder. This hidden layer enables the new model SDAE-DE-ConvMF to learn effective potential factors and to obtain similarities and relationships between users and movies.

Similarly, the potential model for the movie will not be introduced again here, and finally our model calculation formula is as follows:

$$R_{ij} \approx U_i V_j^T = sdae(R_{ij}, Y_i) \left(cnn(W, X_j) + \varepsilon_j \right)^T \quad (2)$$

IV. EXPERIMENTS

A. Dataset

This paper uses the public dataset MovieLens[7], a collection of data collected and organized by a laboratory at the University of Minnesota. For the datasets in MovieLens with many different sizes, we are experimenting with ml-100k, ml-1m, ml-10m and Amazon's AIV datasets. We use the ml-10m, which contains 9,945,875 ratings of 10,073 movies made by 69,878 users.

B. Evaluation Method

The evaluation index used in this paper is the mean square error root RSME, which is directly related to the objective function of the traditional rating prediction model. When the RSME value is smaller, it indicates that the performance of the model is better, and vice versa.

$$RMSE = \sqrt{\frac{\sum_i^U \sum_j^V (R_{ij} - \hat{R}_{ij})^2}{total \ of \ ratings}} \quad (3)$$

Where R_{ij} is the rating of user i on movie j , \hat{R}_{ij} denotes the corresponding predicted rating, U, V represent the number of users and the number of movies, respectively.

C. Results

We compared with based model Probabilistic Matrix Factorization(PMF), and ConvMF. The results of each model on the ml-10m data set are shown in Table 1 below.

TABLE I. RSME OF DIFFERENT MODELS

Model	ml-10m
PMF	0.8311
ConvMF	0.7958
DE-ConvMF	0.7820
SDAE-DE-ConvMF	0.7713

Table 1 shows the RSME in different models, in which we can observe that DE-ConvMF, SDAE-DE-ConvMF achieve higher result than ConvMF, improve 1.73%, 3.08%, respectively. As a result, it demonstrates the domain embedding and user side information are useful to the rating prediction.

V. SUMMARY

In this paper, we proposed two new models, DE-ConvMF and SDAE-DE-ConvMF, respectively. Through the experiments indicate that double embedding architecture use domain embedding for pre-train and SDAE use user side information both improve the rating prediction of accuracy. For the future work, we will use the new model BERT to pre-train our word vector and optimized our model to capture the user interest which will change by time.

REFERENCES

- [1] D. Kim, C. Park, J. Oh, S. Lee, H. Yu, Convolutional matrix factorization for document context-aware recommendation, in: Proceedings of the 10th ACM Conference on Recommender Systems, in: RecSys '16, ACM, New York, NY, USA, 2016, pp. 233–240.
- [2] Ketkar N. Convolutional Neural Networks[J]. 2017.
- [3] Salakhutdinov et al. Probabilistic matrix factorization. NIPS(2008): 1257-1264.
- [4] Wang H , Wang N , Yeung D Y . Collaborative Deep Learning for Recommender Systems[C]// KDD 2015. ACM, 2015.
- [5] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.
- [6] Xu H , Liu B , Shu L , et al. Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction[J]. 2018.
- [7] Information on <https://grouplens.org/datasets/movielens/>