ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0010

A Two-layer Network Recommendation Model Combining Attention and Review Text

Lifen Li[1, 2], Zhao Xin[1]

[1]Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University (Baoding), Baoding, Hebei, 071000, China

[2]Computer Department, North China Electric Power University (Baoding), Hebei, 071000, China

Abstract

With the exponential growth of information on the Internet, the problem of inform-ation overload troubles most netizens, and the recommendation algorithm comes into being. This paper proposes a two-layer neural network recommendation model integrating attentio-n matrix and comment text, which consists of trunk network and target network. The trunk network introduces attention matrix and then uses the target network introduces additional potential vector pairs (target user, target project)using the same convolutional neural networkas the trunk network. The introduction of attention mechanism in the trunk network can bet-ter characterize user preferences and project characteristics, and the influence of user prefer-ences can be effectively reduced by jointly training with the target network. The experime-ntwas conducted on a large public dataset, Amazon-movie, and the results show that this me-thod can effectively improve the recommendation accuracy, and mean squared error (MSE) outperforming other excellent baseline recommendation models.

Keywords

Deep Learning; Recommendation Algorithm; Double-layer Network; Attention.

1. INTRODUCTION

With the improvement of computer computing power and the acceleration of GPU matrix computing, deep learning has gradually become a hot research field, and natural language processing technology has also made breakthroughs [1].Natural language processing technology has also been further paid attention to and utilized,and its advantages in text content mining are very obvious, and it is applied to the recommendation algorithm to analyze the processed comment text,which also provides a new research direction for improving the accuracy of the recommendation algorithm.At present, the deep learning techniques used in recommendation algorithms mainly include convolutional neural network CNN [2],recurrent neural network RNN [3],attention mechanism [4,5], autoencoder [6]and so on.

CNN and RNN networks in deep learning can have good effects when processing text information and retain word order information well.The ConvMF (convolution matrix factorization) model proposed by Kim et al.[7]uses CNN to process comment text information, and combines convolutional neural networks and probability matrix decomposition on this basis to preserve word order in the text.However, the model ignores the user's comment information and only considers the comment text information of the project. Subsequently,Zheng et al[8]proposed the DeepCoNN (deep cooperativeneural network) model, which opened a precedent for using both user and project comment text information, using two parallel convolutional neural networks to process text data, and finally using a factorization machine to score and predict after a layer of coupling, and achieved good results.The Transnets model proposed by Catherine et al. [9] effectively reduces the influence of user preference by introducing additional latent vector expressions (target users, target items) to the training of additional networks, and further improves the accuracy of recommendation algorithms.

Most of the above recommendation algorithms are mainly based on numerical data or some artificial classification data when users interact with the project, and the performance and recommendation accuracy of the recommendation algorithm will also be affected to a certain extent for the limited depth and breadth of feature extraction of the data. Therefore, based on Convolutional Neural Network (CNN) and attention mechanism, this paper proposes a new two-layer neural network recommendation model that integrates attention and comment text to improve the performance and recommendation accuracy of recommendation algorithm. Using MSE [10] as an evaluation index, ablation comparison experiments are carried out on public datasets in five different domains of Amazon,and the results show that the proposed algorithm is superior to a number of excellent models that have been published so far.

2. MODEL NETWORK STRUCTURE

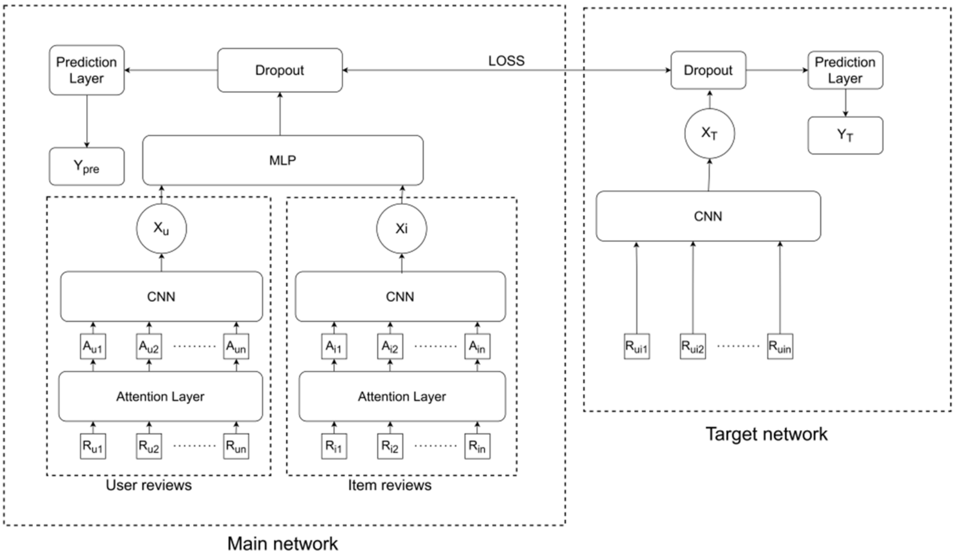
Prediction Layer
Dropout
LOSS
Dropout
Prediction Layer
Y_{pre}
MLP
X_T
Y_T
X_u
Xi
CNN
CNN
CNN
A_{u1} A_{u2} ········ A_{un}
A_{i1} A_{i2} ········ A_{in}
Attention Layer
Attention Layer
R_{u1} R_{u2} ········ R_{un}
R_{i1} R_{i2} ········ R_{in}
R_{ui1} R_{ui2} ········ R_{uin}
User reviews
Item reviews
Target network
Main network

Figure 1. Network structure of two-layer network recommendation model integrating attention and comment text

Based on the deep learning recommendation algorithm model, this paper proposes a two-layer network recommendation model RATN (Review text and Attention Two-layer Network Recommended model) that integrates attention and comment text, and the model structure is shown in Figure 1.The model includes two parallel networks, namely the destination network and the backbone network, the backbone network obtains two preliminary n-dimensional vectors representing the user comment text Run and the project comment text Rin by using the word embedding model [11], and the two n-dimensional feature vectors are extracted through a word vector weight update matrix based on the attention mechanism for feature depth extraction to obtain the word vector sequence, and then through parallel two CNNs with the same structure.The user and the project obtains the deep feature representation Xu, Xi, which is fed into a multilayer perceptron to obtain a final layer representation, and the final score is scored and predicted by a factorization machine [12] (FM). The target network converts the comments (target user, target item) into vector representation XT through the same CNN processing module, and then obtains the processed feature representation through the Dropout layer, and the final score is also given by FM, which uses the user's real review text on the item, similar to a sentiment analysis task. The network structure used to model users and projects in the backbone network is the same, only the inputs are different. Therefore, the paper focuses on the user modeling part of the backbone network (the project modeling process is similar) and the target network modeling process.

2.1. Backbone network modeling

2.1.1 Text Feature Representation Module

Word vector representation is currently one of the most effective text data processing methods in the field of natural language processing, the principle is to use the word embedding model,the comment text of users and items in the data set is mapped to a new expression space through the model function, and a new word vector matrix is obtained that is, f:wordn→Rn, where wordn is each word in the dictionary after processing, and Rn is an n-dimensional feature vector obtained by parameterized function mapping. In this way, the data can be preprocessed to effectively avoid the problem of data sparseness caused by excessive vocabulary, and the data can be reduced in dimensionality.At the same time, the interrelationship between word pairs is added to achieve sentence-level representation.

In the text feature representation module, the comment text is preprocessed by the data and represented as a word embedding matrix. The specific steps for processing dataset comments in this article are as follows:

1)All comment texts are available in all lowercase words through python's NLTK tool. Stop words (the, and, is, etc.)and punctuation marks are considered segregated characters and are retained.

2)Using 5000 words with the highest word frequency in the corpus of the skip-gram model,the pre-trained n-dimensional word embedding vector Mu is:

$$M_{\mathrm{u}} = \{R_{i1}, R_{i2}, \cdots\cdots, \ R_{\mathrm{in}}\} \tag{1}$$

2.1.2 Word vector weight update algorithm based on attention mechanism

This article sets a weight update layerbased on the attention mechanism before the input of the convolutional layer. This layer enables a higher focus on focused content by re-weighting the word vector sequence of the input layer. The design principle of this layer is to use the attention mechanism to update the weight of the original input to produce an input diagram with the same size as the original input, and retain the original input, and expand the input channel of the convolutional layer to dual channels for input

The overall flow of the attention mechanism algorithm designed in this paper is as follows :

Algorithm: Word vector weight update based on attention mechanism

Input: Word vector matrix Mu of all text comment data of user U after data preprocessing,

The respective word vector representation of all words of the user Uall (Wu)

Output: Au, the updated word vector matrix using the attention mechanism

Algorithm steps:

①All word vectors of user U after word embedding processing represent Mu and take out according to a single word.

② (loop 1) uses cosine similarity to calculate the similarity of the word vector representation (Ru(1:n)) of each word in the target user review text and each word in the user comment word vector dictionary (Wu(1:n)Simn), as shown in Equation (2).

$$Sim_n = \sum\nolimits_{i=1}^{n} (R_{u(1:n)}, W_{u(1:n)}) \tag{2}$$

③ (loop 2) normalizes the similarity weights of each word vector through the softmax function, and obtains the attention weight An corresponding to each word vector, as shown in Equation (3).

$$a_n = soft\max(Sim_n) \tag{3}$$

④ On the basis of the word order of the original review text, the attention weights of all word vectors are spliced with them to obtain the attention weight matrix A (Un) of the target user U, as shown in Equation (4).

$$A(U_n) = (a_1, a_2, \cdots\cdots, \ a_n) \tag{4}$$

⑤End loop 1.

⑥End loop 2.

⑦ The attention weight matrix A(Un) is multiplied by the correspondence with the original word vector matrix Mu of user U, and the updated word vector matrix is obtained as shown in Equation (5).

$$A_{un} = A(U_n)M_u \tag{5}$$

2.1.3 Convolutional neural network layer

The embedding layer is followed by the convolutional neural network layer, which takes the word vector matrix obtained by the attention weight update layer as input to obtain the feature vector. Taking user comment text processing as an example, the specific details are as follows:

1) Convolutional layer: Each layer contains m neurons, and the input user word vector matrix Aun is convolved to extract new features. In each neuron, there is a filter filter of size m*t. where m is the feature dimension generated by the word embedding model, and k is the sliding word window size of the design.

In CNN, each convolution kernel is corresponded to a feature map vector mapn by convolution operation, as shown in Equation (6).

$$map_n = f(A_{un} \otimes filter_k + b_k) \tag{6}$$

where f is the activation function ReLU, which is the convolution operation, and bk is the bias corresponding to the filter filterk.

2) Pooling layer: The backbone network takes maximum pooling operations. By selecting the maximum value in the corresponding area of each pooling layer filter, the main feature Oj is finally extracted from the input original feature map vector mapn, and the calculation is shown in Equation (7) to obtain the new feature map vector O, and the calculation is shown in Equation (8) By using the maximum pooling layer, the network compresses the feature map to make its scale smaller, while extracting only the main features, which not only simplifies the complexity of network calculation, but also solves the overfitting phenomenon to a certain extent.

$$o_j = \max\{map_1, \cdots\cdots, map_n\} \tag{7}$$

$$O = \{o_1, \cdots\cdots, o_m\} \tag{8}$$

where m is the number of neurons = number of convolution kernels

3) Connection layer: Enter the feature vector O obtained by the maximum pooling operation into the fully connected layer, multiply it with the weight matrix of the layer and add the bias sum to obtain the classification output Xu, and the calculation formula is shown in (9). The addition of a fully connected layer can further alleviate overfitting and merge for classified output.

$$X_u = f(W * O + b_{all-connected}) \tag{9}$$

2.1.4 Feature fusion score prediction

In this layer, the user features and project features obtained by the text feature extraction module are directly connected to obtain the connection vector Z0.

$$Z_0 = (X_u, X_i) \tag{10}$$

A multilayer perceptron and a dropout layer are used to obtain the final feature representation of the project. This MLP perceptron of an L-layer nonlinear fully connected layer network takes Z0 as input, and the transfer formula from layer l to layer l+1 is as shown in (11). After the calculation of the dropout layer, we finally introduce the Factorization Machine as the evaluator for the corresponding score, and obtain the prediction score Ypre through the calculation formula (12).

$$Z_l = \sigma(Z_{l-1}G_l + g_l) \tag{11}$$

$$Y_{pre} = FM(\delta(Z_L)) \tag{12}$$

where σ is the nonlinear activation function, the weight matrix of the l layer is Gl, and the bias is GL, where δ represents the dropout layer.

2.2. Target network modeling

In the learning of the backbone model, keeping the text information of the target user's comments on the target user in the training set will inadvertently cause the model to rely on potential vector pairs (target user, target item) in the test set, which is impractical. In order to eliminate these irrationalities, a new layer of target networks is created, which is similar to the sentiment analysis problem because it uses actual comments, i.e. a potential vector pair (target user, target item). The target network uses a CNN text processing layer (the same as the backbone network) and a decomposition machine FM to predict the score:

$$X_T = CNN(R_{uin}) \tag{13}$$

$$Y_T = FM(\delta(X_T)) \tag{14}$$

2.3. Two-layer network training method Model

Training adopts two-layer network simultaneous joint training learning to minimize the overall loss, and its training steps are divided into three steps: In the first step, the target network is trained, all its trainable parameters are denoted as θT, and the loss function is the L1 norm between the true score and the predicted score; The second step is to train the backbone network, whose layer before (exclusively) the trainable parameter is denoted as θm, and the loss function is the L2 norm between the expression δ (ZL) output of the dropout layer output and the comment expression XT output of the CNN layer of the target network; In the third step, the remaining trainable parameters of the backbone network are denoted as the L1 norm between the loss function θ0 and the prediction score Ypre.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Dataset

In our experiments, we used Amazon's open-source dataset, which came from 5 different domains to evaluate the proposed model. These datasets come from different types of data from Amazon's website: Digital Music, Toys and Games, Office Product, Instant Video, Kindle Store (DM, TG, OP, IV, KS)). These datasets mainly contain the following: the user's identity, the number of the project, the user's comment information on the project and the rating information (1-5 points). Table 1 provides detailed statistics.

Table 1. Basic information of the dataset

Data set	Userid	Itemid	Number	Sparsity
DM	5,426	3,568	64,475	99.66%
TG	19,412	11,924	167,59	92.75%
OP	4,798	2,419	52,673	99.55%
IV	4,902	1,683	36,486	99.56%
KS	68,223	61,934	982,61	97,66%

In the experiment in this paper, each dataset is divided into training set, test set, and validation set, accounting for 80%, 10%, and 10%, respectively. Among them, the performance of the model is evaluated on the test set, and the verification set is mainly used to adjust the parameters.

3.2. Evaluation indicators

In the experiment, the performance evaluation index of the model selects the mean squared error MSE, and its calculation formula is shown in (14), the smaller the value means that the better the performance of the model, calculated as follows:

$$MSE = \frac{1}{k}\sum\nolimits_{u,i}(R^{'}_{\mathrm{u,i}} - R_{u,i})^{2} \tag{14}$$

where k represents the number of samples, indicates the predicted score, and represents the true score.

3.3. Comparison models

In order to verify the effectiveness of the proposed model, the recommendation model based on the score data is selected for comparison, as follows: MF, PMF, ConvMF [15] DeepCoNN [8] Transnets [9] NAREE[16]

3.4. Analysis of experimental results

Using the Amazon dataset for ablation experiments, the performance of different models was compared on 5 different datasets, and the experimental results are shown in Table 2.

Table 2. Comparison of different model results on Amazon datasets

Model	DM	TG	OP	IV	KS
MF	1.169	1.406	1.787	1.205	1.502
PMF	1.211	1.144	1.092	1.101	0.983
ConvMF	1.083	0.852	0.963	1.199	1.084
DeepCoNN	0.809	0.801	0.739	0.934	0.676
Transnets	0.741	0.712	0.606	0.698	0.612
NAREE	0.793	0.784	0.717	0.793	0,631
(ours)	0.744	0.707	0.589	0.686	0.606

From the comparison of the results of ablation experiments, it can be seen that the performance of the two-layer recommendation model(RATN) that integrates review text and attention matrix is better than that of other models. From the results of Table 2, it can be seen that RATN has achieved better results than other models on different datasets.

4. CONCLUDING REMARKS

Based on the inspiration of the network structure of adversarial generative networks, a two-layer neural network recommendation model RATN integrating attention matrix and comment text is proposed. When processing the scoring matrix of the backbone network, the attention matrix is first used to process the word vector matrix to obtain its deep features, deep feature learning is carried out by deep learning algorithm, and finally the extracted features are fused. The target network introduces additional pairs of potential vectors (target user, target project) to process text extraction features using the same convolutional neural network as the backbone network. The results of ablation experiments show that the proposed RATN model can improve the accuracy of prediction to a certain extent and achieve better results than other models.

REFERENCES

[1] COLLOBERTR, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning[C]Proceedings of the 25th international conference on machine learning(ICML'08).Helsinki,Finland:ACM, 2008:160—167.

[2] ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of research on convolutional neural networks [J]. Chinese Journal of Computers, 2017,40(6):1229-1251.)

[3] GAO Maoting, XU Binyuan. Recommendation algorithm based on recurrent neural network [J]. Computer Engineering,2019,45(8):198-202.)

[4] WANG Yonggui, SHANG Geng. Deep collaborative filtering recommendation algorithm based on attention mechanism[J].Computer Engineering and Applications,2019

[5] AL-SABAHI K.ZUPING Z.NADHER M.A hierarchical structured self - attentive model for extractive document summarization(HSSAS)[J].IEEE Access,2018

[6] WU Y,DUBOIS C, ZHENG A X, et al. Collaborative denoising auto - encoders for top - n recommender systems[C]//Proceedings of the ninth ACM international conference on web search and data mining. California, USA:ACM, 2016:153-162.

[7] KIM D, PARK C, OH J, et al. Convolutional matrix factorization for document context - aware recommendation[C]//Proceedings of the 10th ACM conference on recommender systems. Boston, USA: ACM, 2016:233-240.

[8] ZHENG L, NOROOZI V, YU P S. Joint deep modeling of users and items using reviews for recommendation[C]//Proceedings of the tenth ACM international conference on web search and data mining. Cambridge, UK:ACM, 2017:425-434.

[9] Catherine R ,Cohen W.TransNets:Learning to Transform for Recommendation[C]//the Eleventh ACM Conference.ACM,2017.

[10] ZHAO Jingsheng,SONG Mengxue,GAO Xiang. Review of the development and application of natural language processing[J].Information Technology and Informatization,2019(07):142-145.)

[11] GUO Yuxin,CHEN Xiuhong. Automatic summary model based on BERT word embedding representation and topic information enhancement[J].Computer Science,2022,49(06):313-318.)

[12] JIANG Sheng. Research on intelligent recommendation algorithm based on factor decomposition machine[D].University of Electronic Science and Technology of China, 2021.