Hindawi
Computational Intelligence and Neuroscience
Volume 2020, Article ID 8826557, 10 pages
https://doi.org/10.1155/2020/8826557

Research Article

Online Doctor Recommendation with Convolutional Neural Network and Sparse Inputs

Yongjie Yan[1,2] **Guang Yu**[1] **and Xiangbin Yan**[3]

[1]*School of Management, Harbin Institute of Technology, Harbin 150001, China*
[2]*School of Mathematics and Computer Science, Jiangxi Science & Technology Normal University, Nanchang 330038, China*
[3]*School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China*

Correspondence should be addressed to Guang Yu; yug@hit.edu.cn

Received 10 March 2020; Revised 8 September 2020; Accepted 22 September 2020; Published 15 October 2020

Academic Editor: Giosuè Lo Bosco

The recommendation system in the online medical consultation website is a system to assist patients to find appropriate doctors. Based on the analysis of the current situation of the development of an online medical community (Haodf.com) in China, this paper puts forward recommendation suggestions of finding the right hospital and doctor to promote the rapid integration of Internet technology and traditional medical services. A new recommendation model called Probabilistic Matrix Factorization integrated with Convolutional Neural Network (PMF-CNN) is proposed in the paper. Doctors' data in Haodf.com were used to evaluate the performance of our system. The model improves the performance of medical consultation recommendations by fusing review text and doctor information based on CNN (Convolutional Neural Network). Specifically, CNN is used to learn the feature representation of the review text and the doctors' information. Furthermore, the extended matrix factorization model is exploited to fuse the review information feature and the initial value of the doctors' information for recommendation. As is shown in the experimental results on Haodf.com datasets, the proposed PMF-CNN achieves better recommendation performances than the other state-of-the-art recommendation algorithms. And the recommendation system in an online medical website improves the utilization efficiency of doctors and the balance of public health resources allocation.

1. Introduction

The online medical consultation website is a new type of public health platform formed by the combination of Internet information technology and the medical service industry. There are more and more medical consultation platforms in Chinese online medical website, and people can complete their diagnosis of diseases without leaving their homes. These online medical consultation websites are especially helpful to remote mountainous areas and rural areas that lack higher levels of medical conditions. All kinds of online medical and health services also alleviate the lack of medical resources in certain areas and the imbalance of regional distribution in the country [1]. The analysis of healthcare services based on social media platforms, online doctor reviews, and web-based medical consultation in China are given in [2–4]. However, there are still many problems. For example, the data of each online medical platform in the website are not interoperable, the quality of the platform doctors is uneven, the questions cannot be answered within a limited time, and the condition could be easily misdiagnosed according to the one-sided description.

At present, recommendation algorithms can be generally divided into the following three categories [5, 6], content-based recommendation [7–9], collaborative filtering-based recommendation [10–13], and hybrid recommendation [14–17].

Convolutional Neural Network (CNN) uses layers with convolutional filters to apply local features. CNN was originally invented for computer vision, but, in recent years, lots of studies show that the CNN model has a good effect on recommendation systems and has achieved good results in semantic analysis, search and retrieval, sentence modeling, and other recommendation tasks. Oord et al. [18] directly

use CNN to learn effective representations of songs in a content-based recommendation framework. Kalchblenner et al. [19] described a dynamic CNN, which uses a dynamic k-max pool operator as a nonlinear subsampling function. Their experiments show that the model can achieve good accuracy without external features or other resources. Kim [20] proposes a simple CNN, which uses a single convolution layer and a max pool layer to achieve similar results. He also proved by experience that, using his multichannel model, he achieved 47.4% and 88.1% accuracy in the 5-tag movie comment emotion classification task (sst-1 dataset) and binary classification task (sst-2 dataset), respectively. CNN based relational extraction system is also shown in the following papers [21–23].

The static pretrained word vectors are trained by Mikolov et al. [24, 25] on 100 billion words of Google News. Cicero and Gatti [26] introduced chars CNN by combining word-level embedding and character level embedding as the input of CNN to extract sentence level representation. In the stage of network training, both character level embedding and character level embedding must be trained. They also prove that the structure of feedforward neural network can be effective in sentence sentiment analysis. Lin et al. [27] proposed a real-time and Continua-based Care Guideline Recommendation System (Cagurs) using mobile device platforms. The key idea of Neural Collaborative Filtering (NCF) [28] is to combine MF and MLP with dual-channel structure and learn the user-item interaction using neural networks based on framework for making recommendations. Ma et al. [29] introduce a new group sparse autoencoders algorithm and a new group sparse CNN, which naturally learns the representation of the problem by embedding group sparse self-coding in the traditional CNN.

A lot of research has been carried out on the problem of recommendation, but the existing recommendation algorithms have the following problems:

(1) Most recommendation works associate review information with side information to reduce data sparsity; however, review information has not been fully utilized in related research work. Most of the work of recommendation system using review information focuses on the topic mining of review information using the Latent Dirichlet Allocation (LDA) [30]. However, the model usually uses the word bag model to deal with review information, ignoring the semantic context information of review information. Moreover, when the data is too sparse, the latent feature representation of LDA model learning may not be very effective, and the performance is not satisfactory [31–33].

(2) Most researches on recommendation systems based on machine learning use matrix decomposition technology to recommend items [34]. The method based on the matrix decomposition model is very sensitive to the initialization of the latent feature matrix of users and users' interest points [35]. However, most of the recommendation work based on matrix decomposition avoids or ignores this problem and uses very simple methods (such as random or zero initialization) to initialize the potential features of users and projects.

Because of the above problems, this paper combines the doctors' description document and patients' reviews to mine the potential relationship among doctors, patients, and services and proposes a Probabilistic Matrix Factorization integrated with the Convolutional Neural Networks (PMF-CNN). The main work is as follows:

(1) CNN [32, 36] is used to automatically obtain the deep-seated features in the review information, and the influence of word order and context information on the extracted potential interest features of users can be considered simultaneously to generate a better potential feature representation than LDA model. Particularly when the user reviews matrix is sparse, the use of CNN is helpful to understand the review information in a profound way and generate a better potential model.

(2) This paper proposes an initialization method of hidden layer representation of pretraining data through layer by layer unsupervised learning. We use the depth Stacked Denoising Autoencoders (SDAE) [37] to enter through the review information related to the value of doctors. The best initial value of doctors and patients review information can be obtained by row reconstruction, which can effectively improve the learning efficiency and performance of the matrix decomposition process.

(3) We will test the proposed method on the MovieLens100k dataset and compare it with NCF, which is a de facto benchmark for deep learning recommendation system algorithms, to verify the effectiveness and accuracy of this method. At the same time, we will analyze the important parameters affecting the PMF-CNN recommendation performance. Experimental results show that the proposed algorithm is superior to other advanced algorithms.

(4) This paper proposes a framework based on a deep learning model and probabilistic matrix decomposition model to integrate the relevant information of patients' reviews and doctors' professional knowledge and uses it to predict the patients' preference for the corresponding doctor and gives the specific modeling process. Based on the Haodf dataset, experiments are carried out to verify the performance of the proposed algorithm. Experimental results show that the proposed algorithm is superior to other advanced algorithms.

The paper is arranged as follows: Section 2 introduces the related work, especially in convolutional neural network; Section 3 describes the specific methods of this paper, mainly including the related theory and algorithm implementation; Section 4 is the experiment and analysis; Section 5 gives the summary and outlook.

2. Convolutional Neural Network (CNN)

2.1. The Structure of CNN. Convolutional Neural Network (CNN) is essentially a kind of nonprobabilistic model of multilayer perceptron [38, 39]. However, its architectural differences have significant practical consequences. Although CNN was originally developed for computer vision, its key ideas have been actively applied to information retrieval and natural language processing (NLP), such as search and retrieval, sentence modeling, and classification (traditional NLP tasks). CNN can make good use of the prior information of "spatiotemporal locality" [40] in the data. CNN extracts features from the original data dynamically in the hidden layer between the input layer and the output layer and then applies these features to the classification or fitting of the subsequent output layer. We use CNN to preliminarily explore how to solve the recommendation problem, which has achieved good results in doctor recommendation and demonstrated the feasibility of deep learning in the recommendation system. Sometimes, we want to predict an ordered set (such as a sentence sequence composed of words), such as the emotional tendency of prediction sentences (positive, negative, and neutral). We can find that most of the time, only a few words in a sentence provide useful information, while other words provide little or no information. For example, in the sentence "I am very happy today," the word "happy" has provided enough information to show that the sentence expresses positive emotions. So, the key of the problem is how to select these words with large amount of information. This paper mainly uses CNN to extract this useful information automatically.

There are four layers in the whole CNN network.

(i) Input layer: it is a matrix in which the word vectors corresponding to the words in the sentence are arranged in turn (from top to bottom). If there are n words in the sentence and the dimension of the vector is k, then this matrix is $n \times K$. The type of matrix can be static or dynamic. Static means that the word vector is fixed, while dynamic means that, in the process of model training, the word vector is also regarded as an optimization parameter. For the unknown word vector, it can be filled with 0 or random small positive number.

(ii) Convolution layer: it obtains several feature maps through convolution operation. The size of convolution window is $h \times d$, where h represents the number of vertical words and d represents the dimension of word vector. Through such a large convolution window, several feature maps with 1 column number will be obtained.

(iii) Pooling layer: the pooling layer adopts the method of Max over time pooling. This method simply presents the maximum value from the previous one-dimensional feature map, which represents the most important signal. As you can see, this pooling method can solve the problem of variable length sentence input (because no matter how many values are in the feature map, only the maximum value needs to be extracted). Finally, the output of the pooling layer is the maximum value of each feature map, which is a one-dimensional vector.

(iv) Full connection + softmax layer: the output of one-dimensional vector of pooling layer connects a softmax layer through full connection. The softmax layer can be set according to the needs of the task (usually reflecting the probability distribution on the final category).

The core idea of CNN is to apply a nonlinear function to each word window (k-word window) of the input sentence. This nonlinear function is generally called convolution kernel (called filter in image processing), and this operation is called convolution operation. In this way, the window data of a k-word can be transformed into an m-dimensional vector through the application of filter. In the standard CNN structure, the convolution operation is generally connected with the pooling operation. The most common pooling operations include mean pooling and max pooling. In the field of natural language processing, the maximum pooling is widely used, because through the selecting the maximum value of the features generated by convolution operation is equivalent to obtaining the feature with the largest amount of information, that is, selecting the key words in a sentence. Figure 1 gives a network model architecture of how to perform convolution and pooling operations. In this example, the input sentence is "you should see a doctor today," where the word window K is selected as 3, so there are four window inputs, as shown in the leftmost figure. Suppose that each word is represented by a 2-dimensional vector, so the windows of three words can be represented by a 6-dimensional vector (as shown in the green part of the figure). Convolution operation is equivalent to applying convolution kernel w to each word window, in which m is selected as 3. Therefore, through convolution operation, a 6-dimensional vector will be converted into a 3-dimensional vector. In this example, the data of four windows will be converted into four 3-dimensional vectors, that is, the gray part of the graph. The final pooling operation is to select the maximum value of gray part column and finally generate a three-dimensional vector (blue part in the figure), which is the feature extracted by CNN.

2.1.1. Convolution Operation. CNN subdivides the hidden layer according to the different operation and function and specifically divides it into convolution layer and pooling layer. These two hidden layers can directly learn the information features, so as to extract the features and avoid the extraction of artificial features.

CNN is a model based on multilayer perceptron, but the biggest problem of multilayer perceptron is that it is a fully connected network, so when the input is large, the weight will be especially large. This problem, on one hand, limits the maximum number of neurons that each layer can accommodate and, on the other hand, limits the number of layers of multilayer perceptron, that is, depth. In general, the input needs to be normalized, and the output of each neuron is

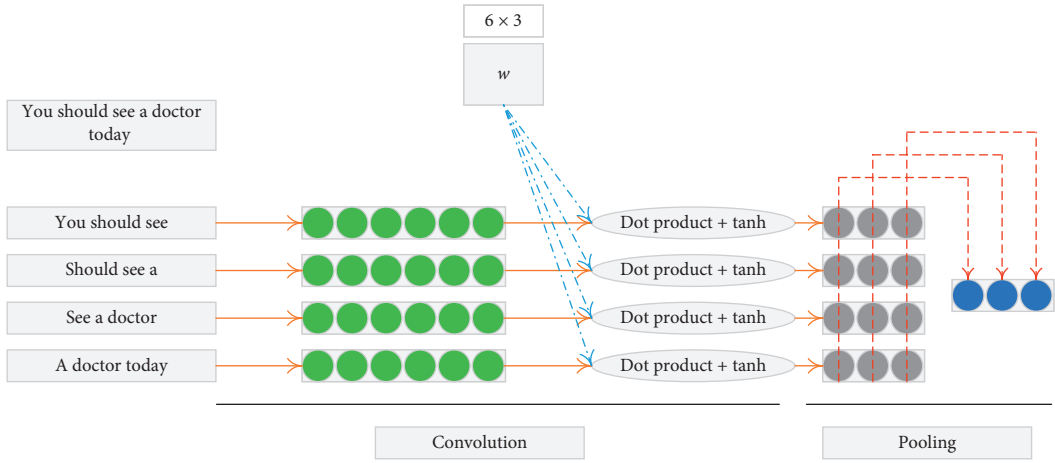

Figure 1: An example of sentence classification of CNN structure.

also normalized under the action of the activation function; in addition, the absolute value of the effective parameters is generally less than 1. In the process of back propagation, multiple numbers less than 1 are multiplied to get smaller values. In other words, with the increase of depth, the residual from the back to the front will be smaller and smaller and even cannot help to update the weights, thus losing the training effect, making the parameters of the front edge layer tend to randomize, and the convolution operation will improve this problem very well.

Convolution operation exists in the convolution layer in the hidden layer. The convolution layer is directly connected with the input layer, and its number is generally consistent with the number of pooling layers. The main function is to enable the artificial neuron to respond to a part of the surrounding units within the coverage, extract the feature directly, and move multiple filters on the input matrix for feature learning. In terms of its structure, the biggest difference between it and the hidden layer in the general artificial neural network is that the connection mode of neurons is not all connected. The operation is similar to full connection, but the operation of full connection layer converts the input into a one-dimensional vector and then performs point multiplication on the one-dimensional vector, while convolution acts on a local area, that is to say, the sensing area of convolution layer is not. It is only part of the neurons in the upper layer. The local information will be integrated into the whole information in the later level. The biggest advantage of this connection is the sharp reduction of the number of weights. Different convolution kernels of different sizes act on the matrix in the middle and will convolute to get different characteristic graphs on the left and right.

The architecture parameter debugging method for sentence classification based on CNN is as follows: Each token of the input sequence is embedded into a 5-dimensional vector, so the input of the model is a 7×5 matrix. The first layer of the model is the convolution layer. There are six convolution kernels in the convolution layer: L1, L2, L3, L4, L5, and L6. Their sizes are 4×5, 3×5, and 2×5. Then, after convolution and activation function, six feature maps with sizes of 4, 4, 5, 5, 6, and 6 are obtained. Six feature maps were obtained by max pooling to obtain 6-dimensional feature vectors. Finally, the corresponding categories are predicted by softmax.

For the input sample $x = \{x_1, x_2, \ldots, x_n\}$, if the word window length is h, then, for the input sample with length N, there are $N - h + 1$ word windows. For ith word window, input $w_i \in R^{(h\,d)}$, which is made up of the word vectors of k words in the window. If the convolution kernel is defined as $W \in R^{(h \cdot d) \times m}$, then the convolution result $P_i \in R^m$ can be defined as shown in the following formula:

$$p_i = f\left(w_i W + b\right), \tag{1}$$

where f is a nonlinear function, such as sigmoid and tanh, while b is an offset.

Every time the window moves, it will perceive the local area covered in the window. The local area that it can perceive is called the perception field. Different window sizes can perceive the characteristics of different area sizes very well. Under the effect of different convolution kernels, we can get different size characteristic graphs. It is worth noting that, before and after each convolution kernel moves, the perceptual weights of each position in the window are the same. This way of weight sharing is also one of the characteristics of convolutional neural network, which helps to reduce the number of weights and the complexity of network model to a large extent.

2.1.2. Pooling Operation. Pool layer is an important part of CNN, which can reduce the problem of overfitting. Its input is from the adjacent upper layer of the convolution layer. Its main function is to further sample the output characteristic map of the convolution layer, that is, to process the extracted characteristics of the convolution layer. The results of the pooling will participate in the subsequent training, so the pooling layer is also called the subsampling layer. Piczak give a good schematic visualization of a typical implementation process of convolution-pooling operation [41].

When the sampling window of 2×2 is used to downsample the information matrix of 4×4, the step size of the

sampling window of the pooling layer is generally the size of the downsampling area. Compared with convolution operation with step size of 1, convolution layer will have overlapping of window areas, while pooling layer generally does not have overlapping of processing areas; that is to say, pooling operation is not continuous. Therefore, pooling operation is more effective in feature reduction when convolution operation and pooling operation are performed with windows of the same size, respectively. In addition, the width and height of the sampling area are not necessarily the same, and the pooled window is not easy to be too large, because the information loss may be serious at the same time of rapid dimensionality reduction.

The features extracted from the convolution layer are regarded as a matrix, and there are two common ways to deal with the matrix in the pooling layer: mean pooling and max pooling. Max pooling in this paper is to take the maximum value of a local area as the feature representative of the area after pooling, comprehensively select the most representative information, and choose the most relevant feature; that is to say, maximum pooling divides the input area into several nonoverlapping subareas, so that each subarea outputs its maximum value. After maximum pooling, the value is calculated according to the following formula:

$$\text{Value} = f\left(W_{pw} \times \max\left(\left(X_p\right)\right) + b\right), \tag{2}$$

where Value represents gray block value, W_{pw} represents pooling weight matrix, X_p represents area matrix covered by pooling window, and b represents offset. f is the activation function.

2.2. Extraction Framework of Medical Consultation Sentence Relationship Based on Neural Network. The task of medical consultation sentence entity relation extraction can be described as follows: given a sentence $s = \{w_1, w_2, e_1, \ldots, w_j, e_2, \ldots, w_n\}$, where e_1 and e_2 are entities, the mapping functions can be defined in the following formula:

$$f(T(s)) = \begin{cases} = 1, & \text{If there is a relationship between } e_1 \text{ and } e_2, \\ = -1, & \text{others,} \end{cases} \tag{3}$$

where $T(s)$ can be regarded as the feature extracted from a sentence containing entity pairs, and the mapping function f determines whether there is a relationship between the two entities, so f can be regarded as a classifier. In this paper, Bayesian classifier can be used as f. It can be seen from Figure 2 that this framework is consistent with our original definition of medical consultation sentence relation extraction in formula (3).

Compared with the traditional relationship extraction system based on machine learning algorithm, the proposed framework uses CNN neural network for automatic feature extraction on the basis of word vector, thus avoiding the process of manual feature extraction. In this paper, multichannel word vectors are introduced as the input of CNN, static and nonstatic pretraining word vectors are used, and the middle feature map is the sum of two feature maps.

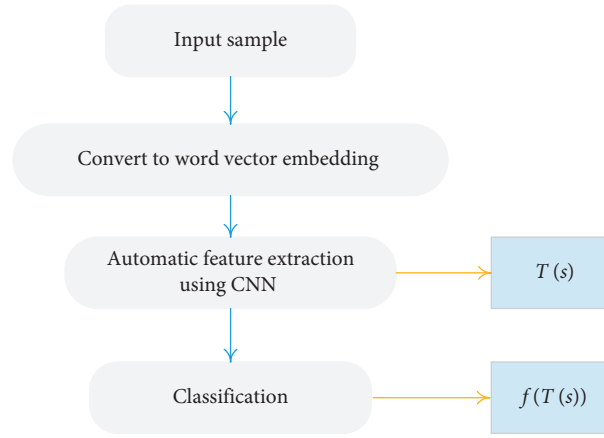

Figure 2: Medical consultation sentence relation extraction via neural networks.

3. Our Approach

In this section, we first describe the definition of the rating prediction task and the notation that we are going to use in this paper. Then, we introduce our CNN model to extract the sequential reviews of patients and doctors. At last, we utilize the sequential features as side information in the feature based collaborative filtering framework to make the final prediction.

3.1. Problem Definition. Given N patients (users) and M doctors (items), the rating r_{ij} is the rating given by i^{th} patient for j^{th} doctor. In the common real-world situations, patients usually rate on a fraction of doctors, not on the whole items. Therefore, those ratings entail a big and sparse matrix $R \in R^{N \times M}$. The goal of recommendation system is to make a prediction on the missing ratings. Based on that, we will know the preference of a patient on the doctors that he/she never rates and recommend high score items to him/her.

Table 1 summarizes the symbols used in the paper. In the next subsection, we will propose a CNN model to extract sequential features of users and items.

3.2. Doctor Recommendation Model. Figure 3 shows a detailed representation of the learning process of each component of the model. The dotted border on the left represents the preprocessing component of patients' reviews information, and the dotted border on the right represents the feature learning component of doctors' categories information. The input is a triple u, i, x, where u represents the user (patient) set, i represents the item (doctor) set, and x represents the reviews information set. Specifically, by learning the initialization parameters of users and items through SDAE, the optimal user characteristics and doctor characteristics are obtained; the potential feature vectors are obtained by Frequency–Inverse Document Frequency (TF-IDF) and learning reviews information through CNN network. Then, by fusing the review feature, the feature of the reasonable doctor ranking is obtained and the score of the doctor is predicted. The following is a detailed introduction

Table 1: Summary of notations.

Notation	Description
N	Number of patients
M	Number of doctors
K	Dimension of latent factors
D	Dimension of sequential features
$R \in R^{N\times M}$	Rating matrix
$U \in R^{N\times K}$	Latent factors of patients
$V \in R^{M\times K}$	Latent factors of doctors
$X \in R^{N\times D}$	Sequential features of patients
$Y \in R^{M\times D}$	Sequential features of doctors

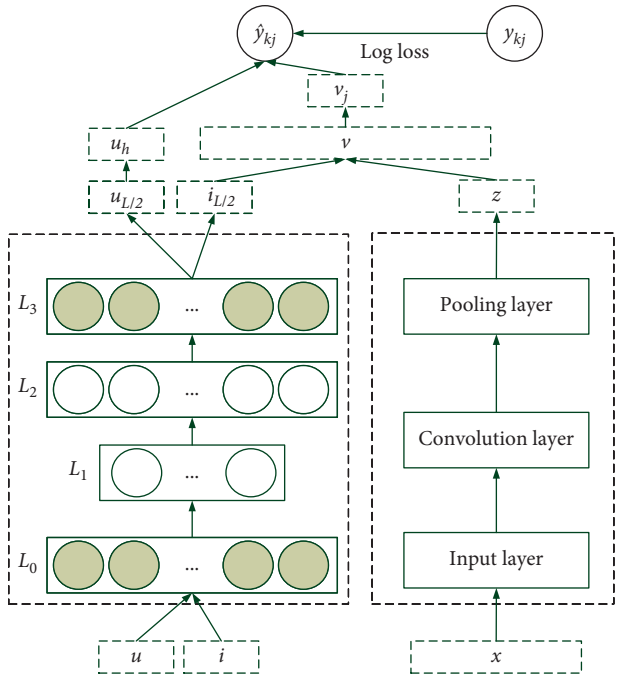

Figure 3: Doctor recommendation model based on probability matrix decomposition of hybrid neural network.

to the learning process of each component of the model in Figure 3.

Table 2 summarizes our baseline PMF-CNN model.

ReLU (corrected linear unit) [42] is an activation function commonly used in deep neural networks. For patients' reviews documents, a simple baseline is to select the nouns with the highest frequency, i.e., appearing in as much reviews as possible. The common approach to extract feature is the term Frequency–Inverse Document Frequency (TF-IDF) as

$$\mathrm{TF-IDF}(x) = \mathrm{TF}(x) \times \mathrm{IDF}(x) = \left(\log \frac{N}{N(x)}\right) \times \left(\log \frac{N(x)}{N(x)+1} + 1\right), \tag{4}$$

where N is the quantity of whole text data and $N(x)$ is the quantity of text data in Table 2.

Probabilistic Matrix Factorization (PMF) is a classic collaborative filtering method to solve this problem [43, 44].

Table 2: PMF-CNN baseline architecture.

Parameter name	Parameter setting	Description
CNN layer	2	The number of CNN layers
HiddenSize	16	The number of hidden layers
Filters	2	The number of filters
KernelSize	4	The number of kernels
Strides	2	The number of strides
Activation	ReLU	Activation function
Max pooling	1	The number of max poolings
Flattened	1	Flattened convolution
Fully connected	1	Fully connected layer
Dropout%	0.10	Discard rate
λ	0.01	Regularization coefficient

It aims to find a K dimensional low rank matrix $\widehat{R} \in R^{N\times M}$ where $\widehat{R} = UV^T$ with $U \in R^{N\times K}$ and $V \in R^{M\times K}$ are two matrices of rank K encoding a dense representation of the patients and doctors with

$$\arg\min_{U,V} \sum_{(i,j)\in K(R)} \left(r_{ij} - \overrightarrow{u}_i^T \overrightarrow{v}_j\right)^2 + \lambda\left(\left\|\overrightarrow{u}_i\right\|_{\mathrm{Fro}}^2 + \left\|\overrightarrow{v}_j\right\|_{\mathrm{Fro}}^2\right), \tag{5}$$

where $K(R)$ is the set of indices of known ratings, $\overrightarrow{u}_i$ and $\overrightarrow{v}_j$ are the corresponding line vectors of U and V, λ is the coefficient that controls the influence of L2 regularization, and $\|\cdot\|_{\mathrm{Fro}}$ is the Frobenius norm.

In Algorithm 1, the pseudocode of PMF-CNN algorithm is given, including training, testing, and prediction.

To train two networks simultaneously using a single loss function in Figure 3, this paper combines the outputs of both networks by concatenation. From here, the interaction of patients' review features with doctor features is done via a PMF in which the details are not provided. However, the goal of the PMF is to capture second order interactions between patients and doctors. The PMF-CNN loss function including PFM:

$$\mathrm{PMF-CNN} = \widehat{\beta}_0 + \sum_{i=1}^{|\widehat{z}|} \widehat{w}_i \widehat{z}_i + \sum_{i=1}^{|\widehat{z}|} \sum_{j=i+1}^{|\widehat{z}|} \langle \widehat{v}_i, t\widehat{v}_j n\widehat{z}_i q\widehat{z}_j \rangle, \tag{6}$$

where $\widehat{\beta}_0$ is the global bias, $\widehat{w}_i$ models strength of i^{th} variable in $\widehat{z}$, and $\langle \widehat{v}_i, \widehat{v}_j \rangle$ is the 2^{nd} order interaction.

4. Experimental Study

4.1. Data Collection. As a new mode of public health service, online medical website has developed rapidly in China. Haodf.com (referred to as Haodf) only includes doctors from public hospitals, but not from private hospitals. When doctors register and open medical services, they need to submit professional title certificates, qualification certificates, and so on. Haodf also has special departments to verify the authenticity of doctors' information, so doctors are all true. Haodf collected 3856035 real votes, comments, and thank-you letters from 194.65 million patients in 605066 doctors'

Input: Interaction matrix between patients' reviews and doctors' categories
Output: The predictive scoring matrix of patients' reviews and doctors' categories
(1) **for** $t < T$ **do**
(2) Select a review matrix X_i randomly from the reviews for training;
(3) the training batch size is β_0, and the size of each batch is B;
(4) calculate the loss L_{cnn} in the training process;
(5) **if** $t > T$ or L_{cnn} is small enough **then**
(6) Use (5), (6), and (9) to calculate the predicted score value of each instance;
(7) Optimizing the parameters of the model by minimizing (3);
(8) **end if**
(9) **end for**

Algorithm 1: PMF-CNN algorithm.

outpatient clinics in 9823 public hospitals across the country. They share their experience of diagnosis and treatment or their subjective feedback on the treatment effect of the doctor, their psychological care for the patient, and their attitude towards coping with the disease together with the patients, which is a good help for more patients to identify their own diseases: which doctor should the patient look for and which doctor is the best that the patients should trust and entrust their own recovery or even life. Relying on the Internet, intellectualization is the inevitable trend of online medical consultation website. Through intellectualization, Haodf can provide more accurate medical and health services, so as to help the construction and development of public health.

The MovieLens100k 1 dataset was used as benchmark experiment dataset. 80% of the dataset is randomly divided as training data and the remaining 20% is used as test data. From 21 July to 20 August 2018, approximately 2 million doctor patient interaction data were obtained from Haodf through web crawler. The data with personal web pages on the online platform of doctors were collected and analyzed from the aspects of medical life, patient visits, and patient satisfaction. The doctors and hospital resources provided by Haodf online platform are mainly public tertiary hospitals, and the professional titles of doctors are mainly intermediate and deputy senior. Only one-third of the doctors with open personal homepage have high activity, and the patients' satisfaction score of the platform is high. The analysis tool of review and ontology used in this paper has two parts: Jieba Chinese word segmentation 2 and Chinese emotional vocabulary ontology of DUTIR 3, respectively.

The density metric [45] in Table 3, which means that how much elements are rated, is calculated according to the following equation:

$$\begin{aligned}\text{Density} &= 100 \times \frac{\#\text{ available ratings}}{\#\text{ all possible ratings}} \\ &= 100 \times \frac{\#\text{ available ratings}}{\#\text{ users} \times \#\text{ items}}.\end{aligned} \tag{7}$$

4.2. Evaluation Criteria. When we evaluate a recommendation system, it is not possible to evaluate only one user's recommendation list and corresponding results, but the entire test set of users and their recommendation list results. The Average Precision (AP) reflects that the indicators are somewhat similar to the concept of recall, except that it is the sequentially sensitive recall. The AP for u is defined as

$$AP = \frac{1}{\left|\mathcal{I}_u^{te}\right|} \sum_{i \in \mathcal{I}_u^{te}} \frac{\sum_{j \in \mathcal{I}_u^{te}} \delta\left(p_{uj} \prec p_{ui}\right) + 1}{p_{ui}}, \tag{8}$$

where p_{ui} indicates the sort position of the item i in the recommendation list. $p_{uj} \prec p_{ui}$ indicates that the item j is in front of the item i in the sort list for user u.

In this paper, the Mean Average Precision (MAP) and the Normalized Discounted Cumulative Gain (NDCG) are used as evaluation indexes to evaluate the performance of recommendation algorithm 1. The MAP indicates the proportion that the first n items recommended can hit the user's actual preference, while the NDCG indicates the ranking quality of the recommendation list.

The MAP is divided into two parts. First, the average accuracy of sorting is calculated, and then the average accuracy of the whole is calculated. The MAP is just the average of all users' AP. The MAP for u is defined as

$$\text{MAP@}u = \frac{\sum_{u \in \text{U}^{te}} \text{AP@}u}{\left|\text{U}^{te}\right|}. \tag{9}$$

Then, the evaluation scores of different users' recommendation lists need to be normalized by the NDCG. The value of the NDCG is between $(0, 1]$. The NDCG@u for u is defined as

$$\text{NDCG@}u = \frac{\sum_{i=1}^{p}\left(2^{\text{rel}_i} - 1/\log_2(i+1)\right)\text{@}u}{\left|\text{U}^{\text{te}}\right|}, \tag{10}$$

where rel_i indicates the relevance of the recommendation results in position i and u indicates the size of the recommendation list to be examined. Then,

$$\text{NDCG@}u = \frac{\sum_{u \in \text{U}^{te}} \text{NDCG}_u\text{@}u}{\left|\text{U}^{te}\right|}. \tag{11}$$

4.3. Experimental Process and Analysis. Among the evaluation criteria, the performance of the model under different parameters is evaluated many times. In the experiment, MAP,

Table 3: Statistics of the two datasets used in this paper.

Dataset	Items	Users	Ratings	Density (%)	User features	Items features
MovieLens 100k	1,682	943	100,000	6.30	Age, gender, and occupation	Genres and year
Haodf	12,000	58,000	220,000	4.10	Doctors' positional titles	State of an illness

Table 4: Experimental evaluation of MAP and NDCG in different dimensions of u, using two metrics.

NCF	$u = 10$	$u = 20$	$u = 30$	$u = 40$	$u = 50$
MAP	0.1069	0.1039	0.0886	0.0875	0.0869
NDCG@3	0.3788	0.3539	0.3393	0.3049	0.2489
NDCG@5	0.4291	0.4151	0.3698	0.3611	0.2979
NDCG@10	0.4630	0.4456	0.4118	0.4020	0.3161
NDCG@20	0.4571	0.4321	0.4159	0.4127	0.3412
PMF-CNN	$u = 10$	$u = 20$	$u = 30$	$u = 40$	$u = 50$
MAP	0.1278	0.1234	0.1072	0.1068	0.1021
NDCG@3	0.4633	0.4339	0.4160	0.3788	0.3042
NDCG@5	0.5072	0.4748	0.4535	0.4152	0.3651
NDCG@10	0.5451	0.5098	0.4908	0.4461	0.3864
NDCG@20	0.5285	0.5079	0.4952	0.4324	0.4178

NDCG@3, NDCG@5, NDCG@10, and NDCG@20 were selected as evaluation indexes. We compared the evaluation indexes of MAP and NDCG between PMF-CNN and NCF [28] in Table 4.

Table 4 shows the MAP and NDCG results of PMF-CNN and NCF in different dimensions. Obviously, PMF-CNN recommendation with content information and adaptive sampling strategy is better than NCF, which shows that content information and convolution sampling strategy have strong feature extraction ability and generalization ability. The results also prove that the improved model proposed in this paper has good feasibility and validity in online doctor recommendation system.

The training execution time of the PMF-CNN and NFC in Table 5.

Many big data technologies are used in the field of natural language processing [46]. At the same time, deep learning is more and more widely used in the field of big data such as syntax analysis, text classification, and sentiment analysis. In this paper, the diagnosis and treatment data analysis uses big data processing technology. We analyze the effect of PMF-CNN on doctors' recommendation of ophthalmology, which is a common medical category in Haodf website. The satisfaction degree of patients with different degree of education to the doctor's diagnosis and treatment results is different. Patients with primary school education had the highest satisfaction with diagnosis and treatment results, while patients with bachelor's degree or above had the lowest satisfaction. We selected 5 diseases (Cataract, Dacryocystitis, Conjunctivitis, Keratitis, and Myopia) from ophthalmology for a case study. Table 6 shows the top 5 recommendation results of ophthalmologists in Shanghai. There are 2479 ophthalmologists in Haodf website. We found that most of the recommended doctors were affiliated to a famous eye hospital, such as Fudan University Affiliated Ophthalm Otolaryngology Hospital; this powerful specialized hospital has skilled doctors. Our recommendation results were validated in Table 6.

Table 5: Results—training execution time comparisons.

	Embedding	Training time
CNN-PMF	**100 dimensions**	**0 hrs 17 min 45 s**
NFC	100 dimensions	1 hr 49 min 52 s

Table 6: A case study of doctor recommendation in ophthalmology.

Diseases	Doctors
Cataract	Xingtao Zhou, Yinghong Ji, You Li, Luo Yi, and Xiaoying Wang
Dacryocystitis	Yan Wang, Lan Gong, Kaiming Su, Jing Li, and Yifei Yuan
Conjunctivitis	Wenqing Zhu, Jiaxu Hong, Hong Liu, Haifeng Qin, and Xinrong Zhou
Keratitis	Zhensheng Gu, Yanjun Hua, Jiaxu Shen, Chunyi Shao, and Peiquan Zhao
Myopia	Peijun Yao, Meiyan Li, Jing Zhao, Jinghui Dai, and Jifang Liang

5. Conclusions

In this paper, a hybrid recommendation algorithm (PMF-CNN) based on deep learning is proposed for doctor recommendation, and an automatic depth encoder is used to learn the initial value of potential eigenvectors of patients' reviews and doctors' professional knowledge in the process of matrix decomposition. PMF-CNN model uses convolutional neural network to learn the context features of review information, so as to extract more accurate feature representation to realize the modeling of review information. For the modeling of interaction between patients' reviews and doctors' professional knowledge in the matrix decomposition model, the best initial value of potential eigenvectors of patients' reviews and doctors' professional knowledge is learned by using the noise reduction automatic encoder to effectively avoid falling into the local optimal solution in the process of matrix decomposition. Finally, the matrix decomposition technology is used to integrate the above two kinds of modeling to provide the patient recommendation service. The verification results on the Haodf dataset show that PMF-CNN is obviously superior to comparative recommendation algorithm.

However, PMF-CNN has the problem of cold start; that is, it can only recommend on the historical doctors and cannot evaluate other new doctors. Therefore, the following research will consider adding features of medical consultation category and patients' reviews to get the representation of category and reviews, so as to solve the problem of cold start and improve the accuracy of recommendation. In the future work, it will be an interesting direction to integrate multiple context

information based on deep learning framework. It is also a should be adaptive on the basis of a results-driven approach in the interface [47].

Data Availability

All the data used in this paper are obtained by Python crawler programming from the HaodF (https://haodf.com), one of the most popular online medical communities in China.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China (71561013, 71774041, and 71531013) and the Humanities and Social Sciences Fund of Universities in Jiangxi Province (JC17221 and JD18083).

References

[1] J. M. Goh, G. D. Gao, and R. Agarwal, "The creation of social value: can an online health community reduce rural-urban health disparities?" *MIS Quarterly*, vol. 40, no. 1, pp. 247–263, 2016.

[2] G. Hu, X. Han, H. Zhou, and Y. Liu, "Public perception on healthcare services: evidence from social media platforms in China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 7, p. 1273, 2019.

[3] H. Hao, "The development of online doctor reviews in China: an analysis of the largest online doctor review website in China," *Journal of Medical Internet Research*, vol. 17, no. 6, p. e134, 2015.

[4] Y. Li, X. Yan, and X. Song, "Provision of paid web-based medical consultation in China: cross-sectional analysis of data from a medical consultation website," *Journal of Medical Internet Research*, vol. 21, no. 6, p. e12126, 2019.

[5] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, "Introduction to recommender systems handbook," in *Introduction to Recommender Systems Handbook*, pp. 11–14, Springer, Boston, MA, USA, 2011.

[6] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[7] P. Lops, M. d. Gemmis, and G. Semeraro, "Content-based recommender systems: state of the art and trends," in *Recommender Systems Handbook*, pp. 73–105, Springer, Boston, MA, USA, 2011.

[8] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web, Methods and Strategies of Web Personalization*, pp. 325–341, Springer, Boston, MA, USA, 2007.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web Hong Kong*, pp. 285–295, Hong Kong, May 2001.

[10] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work ACM*, pp. 175–186, Chapel Hill, CA, USA, October 1994.

[11] J. B. Schafer, D. Frankowski, J. Herlocker et al., "Collaborative filtering recommender systems," in *The Adaptive Web, Methods and Strategies of Web Personalization*, pp. 291–324, Springer, Berlin, Germany, 2007.

[12] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative filtering recommender systems," *Foundations and Trends in Human-Computer Interaction*, vol. 4, no. 2, pp. 81–173, 2011.

[13] X. Y. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 2009, Article ID 421425, 19 pages, 2009.

[14] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[15] R. Burke, "Hybrid web recommender systems," in *The Adaptive Web, Methods and Strategies of Web Personalization*, pp. 377–408, Springer, Berlin, Germany, 2007.

[16] M. Balabanović and Y. F. Shoham, "Content-based collaborative recommendation," *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[17] C. A. Gomez-Uribe and N. Hunt, "The netflix recommender system: algorithms, business value, and innovation," *ACM Transactions on Management Information Systems*, vol. 6, no. 4, p. 13, 2015.

[18] A. d. V. Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS 2013)*, pp. 2643–2651, Lake Tahoe, NV, USA, December 2013.

[19] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, https://arxiv.org/abs/1404.2188.

[20] Y. Kim, "Convolutional neural networks for sentence classification," 2014, https://arxiv.org/abs/1408.5882.

[21] Z. Yang, N. Tang, X. Zhang, H. Lin, Y. Li, and Z. Yang, "Multiple kernel learning in protein-protein interaction extraction from biomedical literature," *Artificial Intelligence in Medicine*, vol. 51, no. 3, pp. 163–173, 2011.

[22] S. P. Choi and S. H. Myaeng, "Simplicity is better: revisiting single kernel PPI extraction," in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 206–214, Beijing, China, August 2010.

[23] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction," in *HLT/EMNLP 2005, Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 724–731, Vancouver, Canada, October 2005.

[24] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, https://arxiv.org/abs/1301.3781.

[25] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," 2013, https://arxiv.org/abs/1310.4546.

[26] S. D. Cícero and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, Dublin, Ireland, August 2014.

[27] Y.-F. Lin, H.-H. Shie, Y.-C. Yang, and V. Tseng, "Design of a real-time and continua-based framework for care guideline recommendations," *International Journal of Environmental Research and Public Health*, vol. 11, no. 4, pp. 4262–4279, 2014.

[28] X. N. He, L. Z. Liao, H. W. Zhang, and L. Q. Nie, "Neural collaborative filtering," in *Proceedings of the 26th*

International Conference on World Wide Web, pp. 173–182, Perth, Australia, April 2017.

[29] M. Ma, L. Huang, B. Xiang et al., "Group sparse CNNs for question classification with answer sets," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 335–340, Vancouver, Canada, July 2017.

[30] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 6, pp. 993–1022, 2003.

[31] H. Wang, N. Y. Wang, and D. Y. Yeung, "Collaborative deep learning for recommender systems," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1235–1244, Sydney, Australia, August 2015.

[32] D. Kim, C. Parkm, J. Oh et al., "Convolutional matrix factorization for document context-aware recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 233–240, Boston, MA, USA, September 2016.

[33] X. Dong, L. Yu, Z. H. Wu et al., "A hybrid collaborative filtering model with deep structure for recommender systems," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 1309–1315, San Francisco, CA, USA, February 2017.

[34] R. Gao, J. Li, X. Li, C. Song, and Y. Zhou, "A personalized point-of-interest recommendation model via fusion of geo-social information," *Neurocomputing*, vol. 273, pp. 159–170, 2018.

[35] R. Zdunek, "Initialization of nonnegative matrix factorization with vertices of convex polytope," in *Proceedings of the 11st International Conference on Artificial Intelligence and Soft Computing Zakopane*, pp. 448–455, Zakopane, Poland, April 2012.

[36] C. Yang, M. S. Sun, W. X. Zhao, L. Zhiyuan, and E. Y. Chang, "A neural network approach to jointly modeling social networks and mobile trajectories," *ACM Transactions on Information Systems (TOIS)*, vol. 35, no. 4, pp. 1–28, 2017.

[37] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[38] B. T. Hu, Z. D. Lu, H. Li, and Q. C. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proceedings of the Conference and Workshop on Neural Information Processing Systems Montreal*, pp. 2042–2050, Montreal, Canada, December 2014.

[39] Y. Zhang and B. C. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," 2015, https://arxiv.org/abs/1510.03820.

[40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[41] P. Karol, "Environmental sound classification with convolutional neural networks," in *Proceedings of the IEEE 25th International Workshop on Machine Learning for Signal Processing*, pp. 1–6, Boston, MA, USA, September 2015.

[42] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the International Conference on Machine Learning*, pp. 807–814, Madison, WS, USA, June 2010.

[43] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[44] J. Bennett and S. Lanning, "The netflix prize," in *Proceedings of KDD Cup and Workshop 2007*, pp. 3–6, San Jose, CA, USA, 2007.

[45] S. Dooms, A. Bellogín, T. D. Pessemier, and L. Martens, "A framework for dataset benchmarking and its application to a new movie rating dataset," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 3, pp. 1–28, 2016.

[46] V. N. Gudivada, D. Rao, and V. V. Raghavan, "Big data driven natural language processing research and applications," *Handbook of Statistics*, vol. 33, pp. 203–238, 2015.

[47] O. Gambino, L. Rundo, V. Cannella, S. Vitabile, and R. Pirrone, "A framework for data-driven adaptive GUI generation based on DICOM," *Journal of Biomedical Informatics*, vol. 88, pp. 37–52, 2018.