A Study on Doctor Recommendation Model in Medical Guidance Services

Hao Chen[1*], *Wenye Gu*[2], *Haifei Zhang*[1], *Qiong Zhang*[1], *Xiaowei Ding*[3]

[1] *School of Computer and Information Engineering, Nantong Institute of Technology, China*
[2] *Affiliated Hospital of Nantong University, China*
[3] *Jiangsu Chaoshu Information Technology Co., Ltd., China*
chenhao@ntit.edu.cn, fengwu0796@outlook.com, 20160063@ntit.edu.cn, 18862928127@163.com, ntdxw@163.com

Abstract

In medical guidance services, it is of great significance to match the appropriate department and doctor by digging deeper into patients' demands. However, accurate matching of doctors requires the ability to locate the exact department based on the text of the patients' chief complaint, and then select the matching doctor by considering the patients' condition, the doctor's professionalism, and the patients' preference. To this end, this paper proposes a department classification model on the basis of Convolutional Neural Networks (CNN) as well as Robustly optimized BERT approach (RoBERTa) with an attention mechanism. The model firstly extracts the patients' chief complaint texts features by convolution layer, and then introduces the attention mechanism to assign different weights to different features. Subsequently, these features are fused with the features extracted by RoBERTa for classification. In addition, this paper proposes a doctor recommendation algorithm that considers both patient similarity and patient preference. Through the in-depth analysis on the patients' condition claims, various weights are assigned to various influencing factors, and then the matching degree is calculated to achieve the accurate recommendation of doctors. The experimental results reveal that the proposed department classification model's accuracy on the dataset is 93.4%, and the Normalized Discounted Cumulative Gain (NDCG) of the doctor recommended algorithm is 90.7%. In this way, the proposed model effectively improves patient-doctoral matching with excellent performance.

Keywords: CNN, RoBERTa, Attention mechanism, Doctor recommendation, NDGG

1 Introduction

As a significant component of the medical service system, the medical guidance service can coordinate and solve problems encountered by patients when they visit the clinic, providing patients with convenient and quick services [1]. Through the medical guidance service platform for patients to match the appropriate departments and doctors, the rational use of medical resources can be realized. However, in the traditional medical guidance services, there is a problem that the patients are in lack of professional knowledge, which leads to the blindness in medical selection. As a result, it is of great significance for medical guidance services to combine emerging information technology with traditional medical services, obtain useful medical information according to the main complaints of patients, and match appropriate departments and doctors to improve the efficiency of medical care for patients [2].

In medical guidance services, patients expect to quickly select the right doctor according to their own conditions as well as their own preferences for doctors, which can reduce the misdiagnosis rate of diseases and save time for medical treatment. Therefore, this paper combines the patients' chief complaint texts and the patient-doctor relationship. Firstly, the appropriate department is selected based on the patients' chief main complaint texts. Then, the two aspects (patient similarity and patient preference) are combined to design an accurate doctor recommendation model.

Since the patients describe the symptoms of their diseases in the form of chief complaints, the department to be visited can be selected by extracting the features in the text of the patients' chief complaint. Thus, in this paper, the study of the department recommendation problem is transformed into a multi-classification problem to deal with the patients' chief complaint texts [3]. The study of department recommendation mainly contains rule-based or statistical machine learning methods, and deep learning-based methods. Rule-based or statistical machine learning methods are based on the medical professional knowledge base, using certain rules for text feature extraction, and then classifying the text [4]. Botsis et al. [5] adopted a rule and machine learning approach to achieve classification of vaccine adverse report texts. Krämer et al. [6] proposed a supervised machine learning approach based on patients' diagnostic texts so as to achieve classification between emergency departments and general departments. The above-listed methods demonstrate the effectiveness of classification in medical texts. However, rule-based methods require manual formulation by experts, which leads to computational inefficiency. Traditional machine learning models require manual feature extraction together with fine feature engineering to ensure classification accuracy. With the development of technology, deep learning methods began to be applied to medical text classification [7]. Deep learning methods mainly represent high dimensional

*Corresponding Author: Hao Chen; E-mail: chenhao@ntit.edu.cn
DOI: https://doi.org/10.70003/160792642024092505002

sparse raw data as low dense feature vectors, which can automatically learn some connections between data and labels [8]. According to some studies, deep learning methods are far superior to machine learning in text classification [9]. Kim [10] first applied Convolutional Neural Networks (CNN) to sentence-level classification tasks, and verified the effectiveness of CNN in extracting text features. However, CNN's perceptual field of view is localized, and multiple convolutions are required to expand the perceptual field of view, resulting in duplicated extracted features. Through the introduction of an attention mechanism [11], it is possible to focus the limited attention on the focal information so that the extracted features are more critical. It has been indicated that adding the attention mechanism to CNN can further improve the model's classification performance [12]. However, Due to the fact that CNN primarily focuses on local features, the CNN effect is more noticeable in short text classification. For long texts, the text content is long and complex, and the utilization of CNN to extract long text features is likely to cause information loss. Long Short-Term Memory (LSTM) can better capture the dependency relationship in the text over a long distance [13]. However, when the text content is related before and after, it cannot effectively capture the bi-directional dependency between the front and the back. For finer-grained classification tasks, Bi-directional Long Short-Term Memory (BiLSTM) can better capture bidirectional semantic relationships in context [14]. However, the application of Robustly optimized BERT approach (RoBERTa) [15] enables the dynamic semantic representation of words, which can better learn the contextual information of the text. Because the length of patients' chief complaint texts varies, the text classification model can be designed by combining the advantages of the previous methods in order to effectively extract the features of long texts and short texts simultaneously. Models mixing CNN and BiLSTM and combining the attention mechanism were verified to have better classification results [16-18]. Models based on the mixture of CNN and RoBERTa have also achieved high accuracy in classification tasks [19-20]. Different models are applied to different scenarios. Based on the characteristics of the patient's chief complaint texts, this paper combines the advantages of CNN and RoBERTa in extracting of text features, introduces the attention mechanism, and assigns different features to different features, so as to realize the classification of departments.

After locating all the doctors in the department, this paper proposes a doctor recommendation model by means of considering both patient similarity and patient preference. This model matches patients with appropriate doctors, aiming to reduce the pressure on medical resources and improve patient satisfaction. Relevant studies have indicated that multifaceted consideration of the patient-doctor relationship is capable to improve the accuracy of doctor recommendations [21-22]. Most studies in the field of traditional doctor recommendation have utilized the patients' chief complaint texts as the main discussion factor [23]. Narducci et al. [24] proposed to calculate similarities between patients, and then compare based on the health data shared by the community to generate a list of suitable doctors. Kou [25] proposed to compute the similarity between patients and doctor's' history patients based on Word2vec, Term Frequency–Inverse Document Frequency (TF-IDF), and cosine similarity to achieve doctors' recommendations. Yan et al. [26] extracted the features of review information and doctor information through CNN, and then established patient and doctor matching through matrix decomposition technology. Zheng et al. [27] designed a dialogue-based doctor recommendation model by combining patient information, patients' chief complaint contents, and doctor-patient dialogue information. The above approaches consider the fact that the more similar a patient's condition is, the more accurate the patient's matching doctor is. However, in the actual medical guidance service, the patient's choice of doctor is also influenced by the doctor's professionalism and medical resources. If all the patients select the chief physician, it is likely to lead to an insufficient number of doctors, resulting in a strain on medical resources. Furthermore, a doctor's professional competence cannot be evaluated solely on the basis of his or her title, which is related to the number of years the doctor has worked as well as the evaluation of the doctor by historical patients. Therefore, it is necessary to make a comprehensive decision based on the full consideration of the similarity of patients' conditions, the pressure of medical resources, and patients' satisfaction degree. Yang et al. [28] proposed a systematic decision support doctor recommendation model based on the combination of two factors including patient preference and online reviews to improve patient visit satisfaction and doctor recommendation accuracy. Singh et al. [29] incorporated features such as doctor experience, doctor ratings, and communication skills into the designed model, then computed the skill scores of each doctor, and finally implemented doctor recommendation by applying K-Nearest Neighbor (KNN) to achieve doctor recommendation. By exploring the doctor recommendation method through a variety of factors, we are able to better match patients with doctors by taking into account hospital medical resources and patient satisfaction while ensuring the accuracy of doctor recommendation. However, the weight of various influencing factors on the accuracy of the doctors' recommendations is not the same. In this paper, we design a doctor recommendation model with the objectives of matching patients' conditions, balancing medical resources and improving patient satisfaction, assign different weights to different influencing factors, and compare the recommendation effect with different doctor recommendation models. The main contributions of this paper are as follows:

(1) A deep learning-based department classification model is proposed. Firstly, local features of texts are extracted by CNN, and then the attention mechanism is introduced to extract key features. The text context features are captured by the pre-trained model RoBERTa. Secondly, the features extracted from the two sections are fused. Finally, the departmental classification is completed.

(2) Six factors are considered in the doctor recommendation model as follows: text similarity of patients, age similarity of patients, gender similarity of patients,

doctor's working years, doctor's title, and doctor's score. Furthermore, different weights are assigned to different influencing factors to achieve accurate recommendation by doctors. We compared the designed model with other models, and the experimental results verified the effectiveness of the doctor recommendation model proposed in this paper.

The study in this paper is capable to solve the problem of the blind medical treatment caused by the asymmetry of doctor-patient information when patients select doctors, which can reduce the pressure on doctors' resources and improve patients' satisfaction, laying a theoretical foundation for the study on intelligent guidance in medical guidance services. The rest of the paper is organized as follows: Section 2 specifies the proposed methodology as well as the theoretical background; Section 3 presents the results of the experimental study; Section 4 discusses the conclusions and future work.

2 Methods

The overall framework for the design of this paper is shown in Figure 1. The patient's chief complaint texts are input into the department recommendation model in order to match the appropriate department, and thus locate the departmental doctors. The patient similarity and patient preference are employed as inputs to the doctor recommendation model. Consequently, a list of recommended doctors is calculated.

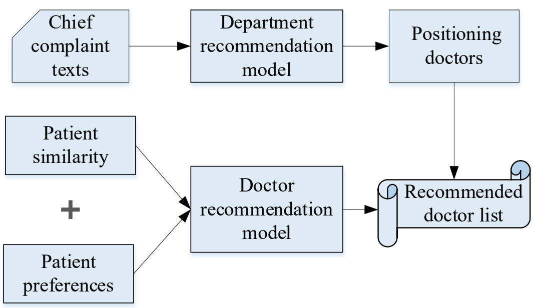


Figure 1. Overall framework of this paper

2.1 Department Recommendation Model

This paper proposes a fusion of CNN-Attention and RoBERTa for department classification model. The structure of the hospital department classification model is shown in Figure 2. The input layer contains RoBERTa-Input layer and CNN-Input layer. The text feature extraction layer contains RoBERTa layer and CNN-Attention layer. First, the patient's chief complaint texts were pre-trained with the models RoBERTa and Word2vec to achieve a word vector representation. Subsequently, text features are extracted, and features are fused through utilizing RoBERTa and CNN-Attention models, respectively. Finally, the department prediction is achieved by Dropout layer, Dense layer, and Softmax layers. Details of these steps are described from Sections 2.1.1 to Section 2.1.5.

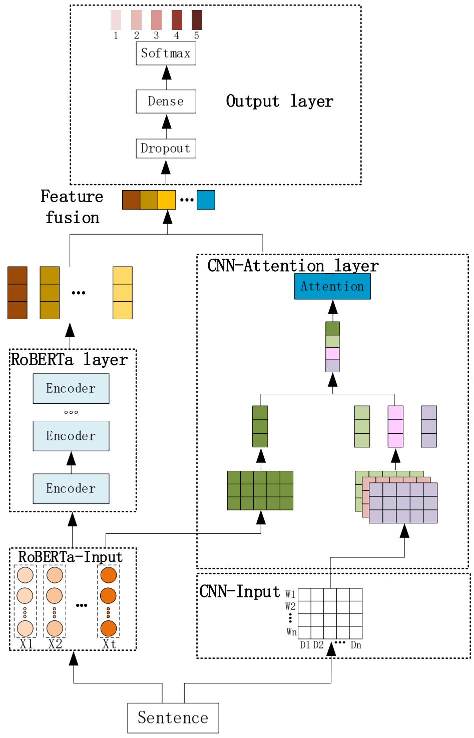


Figure 2. Structure of the hospital department classification model

2.1.1 CNN-Input Layer

Word2vec is a shallow neural network model, whose goal is to convert words into their corresponding vector representations. Word2vec includes the following two important models: the Continuous Bag-of-Words (CBOW) Model and the Skip-Gram model. They are both text vector representations. The CBOW model predicts the central word utilizing adjacent words, followed by the central word to predict the results. The Skip-Gram model is the opposite of the CBOW model, which adopts central words to predict adjacent words. This paper employs the CBOW model.

The CBOW [25] model includes the following three layers: an input layer, a hidden layer, and an output layer. The input layer is a one-hot encoding of contextual words, and the output obtained through the matrix W_{V*N} can be expressed as follows:

$$h = W_{V*N}\left(\sum_{i=1}^{T} X_i\right) \tag{1}$$

Where, T is the window size; V is the dimension of the word, and h is the input of the hidden layer. The value after a further weighted average can be expressed as follows:

$$h' = \frac{h}{T} \tag{2}$$

Where, h' is the output of the hidden layer. After passing through the weight matrix W'_{N*V}, the output can be expressed as follows:

$$m = h' * W'_{N*V} \tag{3}$$

Where, m is processed by Softmax to obtain a vector of l*V. The final result can be expressed as follows:

$$Y_{T,j} = p(W_{y,j} \mid W_1, W_2, ..., W_T) \tag{4}$$

Where, $Y_{T,j}$ is the index with the highest probability in the vector, and the represented word is the middle word of the prediction.

2.1.2 RoBERTa-Input Layer

RoBERTa is improved on the basis of BERT by fine-tuning the masking strategy and data of the model. The Token Embedding, Position Embedding, and Segment Embedding of the words are constructed. The sum of the three vectors is utilized as the input in the model. Token Embedding is to convert each word into a fixed dimensional vector; Position Embedding is to learn a vector representation at each position to represent the text sequence information; Segment Embedding is mainly utilized for sentence pair classification task to distinguish two sentences in a sentence pair.

2.1.3 CNN-Attention Layer

CNN-Attention indicates that the attention mechanism has been introduced after CNN. The text is first converted into a corresponding text vector by Word2vec. Then, convolution operations are performed on the text vectors using convolution kernels in Sizes 3, 4, and 5 respectively. Finally, the fusion is performed after passing the maximum pooling operation respectively. To obtain richer semantic information, this paper also uses RoBERTa's pre-trained text word vector expressions as the CNN layer's input. After that, the features are fused with the features extracted by the CNN layer after Word2vec pre-training. Finally, the result of feature fusion is employed as the input of attention to assign feature weights and focus attention on the key features. The role of the attention mechanism is to be able to correlate global features and assign various weights to various words. The introduction of the attention mechanism in the text classification task can reflect the importance of words in the text and improve the text feature extraction effect.

2.1.4 RoBERTa Layer

RoBERTa is improved on the basis of BERT by fine-tuning the masking strategy and data of the model. The text is fed into the RoBERTa model to obtain the text's corresponding high-order feature vectors. Subsequently, a forward neural network and a Softmax layer are added to achieve text classification. RoBERTa employs a multi-layered Transformer bidirectional encoder as a text feature extractor to achieve bidirectional association between word embedding vectors.

2.1.5 Output Layer

The features extracted by CNN-Attention and RoBERTa respectively are entered into fusion. The fused features are used as the input of the Output layer. The fully-connected layer is then constructed based on the input of the Output layer and the internal medicine department category. Due to the small amount of training data, Dropout layer is introduced to prevent overfitting and improve the model effect. Eventually, the final recommended department classification results are obtained by Softmax.

2.2 Doctor Recommendation Model

The computational flow of this paper's doctor recommendation model is shown in Figure 3. The three factors (gender, age, and text similarity) between the target patient and the doctor's historical patient visits are combined as a patient similarity, and the other three factors (doctor's working years, doctor's title, and doctor's score) are combined as a patient preference. Different weights are then assigned to the above six factors to construct the precise doctor recommendation model in this paper. Details of these steps are described from Sections 2.2.1 to Section 2.2.3.

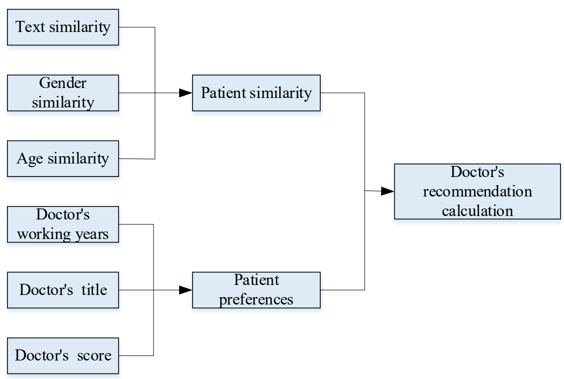


Figure 3. The computational flow of doctor recommendation mode

2.2.1 Patient Similarity

Patient similarity combines the following three aspects: text similarity, age similarity and gender similarity. The Chinese text is divided into words first, then deactivated, and then vectorized. Finally, using the cosine theorem [25], we calculate the degree of similarity between the two texts.

The text similarity formula for patient M and patient N is expressed as follows:

$$\text{TextSim}(M, N) = \text{COS}(\theta) = \frac{\sum_{i=0}^{n-1}(A_i * B_i)}{\sqrt{\sum_{i=0}^{n-1}(A_i)^2} * \sqrt{\sum_{i=0}^{n-1}(B_i)^2}} \tag{5}$$

Where, A and B are both n-dimensional vectors and θ is the angle between Vector A and Vector B. The closer the cosine value is closer to one, the more similar the two vectors A and B are.

Patient gender similarity is also calculated using the cosine theorem, with males represented by the vector [0,1] and females represented by the vector [1,0]. Gender similarity is calculated using the following formula:

$$\text{SexSim}(M, N) = \frac{\sum_{i=0}^{1}(C_i * D_i)}{\sqrt{\sum_{i=0}^{1}(C_i)^2} * \sqrt{\sum_{i=0}^{1}(D_i)^2}} \tag{6}$$

Where, C and D are both 2-dimensional vectors. The correlation between two patients based on their ages is referred to as age similarity. The age similarity is calculated as follows:

$$\mathrm{AgeSim}(M,N)=\begin{cases}1, & |\mathrm{Age_M}-\mathrm{Age_N}|\le 5\\ 1-\dfrac{|\mathrm{Age_M}-\mathrm{Age_N}|}{\varepsilon}, & 5<|\mathrm{Age_M}-\mathrm{Age_N}|\le 15\\ 0, & |\mathrm{Age_M}-\mathrm{Age_N}|>15\end{cases} \tag{7}$$

Where, ε is the age gap parameter between the two patients, which takes the value 25. $\mathrm{Age_M}$ is the age of patient M and $\mathrm{Age_N}$ is the age of patient N. When the age difference between the two patients is less than 5, the age similarity is 1; when the age difference between the two patients is more than 15, the age similarity is 0.

In summary, the formula for calculating patient similarity is as follows:

$$\mathrm{Sim}_\alpha(M,N)=\alpha_1\mathrm{TextSim}(M,N)+\alpha_2\mathrm{SexSim}(M,N)+\alpha_3\mathrm{AgeSim}(M,N) \tag{8}$$

Where, α_1, α_2, and α_3 are the influence factors for text similarity, gender similarity and age similarity. Values are between (0, 1) and $\alpha_1+\alpha_2+\alpha_3=1$.

2.2.2 Patient Preference Setting

Patient preference takes into account the following three factors: working years, title, and score of the doctor. The weight of a doctor's working years is expressed as the ratio of one doctor's working years to the maximum working years among all doctors. It is calculated as follows:

$$\mathrm{WorkRate_t}=\frac{\mathrm{T_t}}{\max\{\mathrm{T_1},\mathrm{T_2},\ldots,\mathrm{T_t},\ldots,\mathrm{T_n}\}} \tag{9}$$

Where, t is the doctor's index, and $\mathrm{T_t}$ is the doctor's working years

The titles of hospital outpatient doctors can be divided into chief physician, associate chief physician, and attending physician. In order to alleviate the pressure on medical resources and balance the fairness of the influence of doctors' titles on their recommendation results, this paper sets the weight of the title of chief physician to 0.5, the weight of the title of associate chief physician to 0.3, and the weight of the title of attending physician to 0.2.

The doctor's score is then expressed in terms of the ratio of positive reviews to all reviews, and the weight of the rating is calculated as follows:

$$\mathrm{FavRate_t}=\frac{\mathrm{X_t}}{\mathrm{Y_t}} \tag{10}$$

Where, X is the number of positive reviews for that doctor, and Y is the total number of reviews for that doctor.

In summary, the formula for calculating patient preference is as follows:

$$\mathrm{Sim}_\beta(t)=\beta_1\mathrm{WorkRate_t}+\beta_2\mathrm{PosRate_t}+\beta_3\mathrm{FavRate_t} \tag{11}$$

Where, $\mathrm{PosRate_t}$ is the doctor t's title weight. β_1, β_2, and β_3 are the influence factors of doctor's working years weight, doctor's title weight and doctor's score weight, with values between (0, 1) and $\beta_1+\beta_2+\beta_3=1$.

2.2.3 Doctor Recommendation Model Calculation

We assume that the target patient is M and patient N is the historical patient seen by the doctor t. In this paper, the formula for calculating doctor t's recommendation similarity is then developed as follows:

$$\mathrm{Sim}(M,N,t)=\alpha\mathrm{Sim}_\alpha(M,N)+\beta\mathrm{Sim}_\beta(t) \tag{12}$$

Where, α and β is the influence factors of patient similarity and preference, respectively, and $\alpha+\beta=1$. Assuming that the number of patients seen by the doctor t is J, we express the average similarity of the doctors' recommendations as follows:

$$\mathrm{AveSim_t}=\frac{\sum_{i=1}^{J}\mathrm{Sim}(\mathrm{U_m},\mathrm{U_i},t)}{J} \tag{13}$$

The above equation represents the recommended doctor t's similarity calculation for the target patient $\mathrm{U_m}$ and the doctor's medical history $\mathrm{U_i}$.

3 Experimental Study and Analysis

This section begins with a description of the data set applied for the experiments and the indicators applied to evaluate the experiments. Subsequently, the hyper-parameters of the department classification model and the doctor recommendation model are experimentally set. Consequently, the superiority of the models is verified experimentally. These details are presented from Sections 3.1 to Sections 3.4.

3.1 Datasets

The dataset for the departmental classification model is sourced from Partner Hospital and is calibrated by professionals to ensure the accuracy of the data. The dataset includes a description of the patient's symptoms as well as the name of the department visited. This paper selects four Internal Medicine Departments. Plentiful examples from the medical guide text data are as revealed in Table 1. The selected dataset contains 8,000 pieces of data. Eighty percent of these were utilized as the training set and 20% as the test set.

Table 1. Example of dataset

Number	Text	Level 1 department	Level 2 department
1	Headache and fever, body sweats easily.	Internal Medicine	Respiratory medicine
2	Men's hair is prone to oiliness, accompanied by hair fall.	Internal Medicine	Endocrinology
3	Stabbing pains in both temples of the head several times a day.	Internal Medicine	Neurology
4	Half an hour after a meal, there is pain in the epigastrium.	Internal Medicine	Gastroenterology

The doctor recommendation model adopts basic doctor information and basic patient information. In this paper, doctor and patient information is obtained from Partner Hospital to create a pool of information, which can be used to recommend doctors to patients during consultations. The pool of information resources includes the doctor's age, gender, title, working years, department, score, the patient's chief complaint texts, the patient's age and the patient's gender. The doctor's historical patient's chief complaints are employed for similarity calculation with the target patient's chief complaints. The patient's age and the patient's gender are applied for similarity calculations with the basic information about the target patient.

3.2 Evaluation Metrics

The department classification model applies accuracy and F1-score as evaluation metrics for classification effectiveness. The accuracy rate is the ratio of correctly classified samples to the total sample. The F1-score represents the summed average of precision and recall. Assuming that the categories of the four internal medicine departments in the dataset are denoted as A1, A2, A3, and A4, we can express the calculation of the accuracy [10] as follows:

$$P_i = \frac{Q_{A_i}}{H_{A_i}} \tag{14}$$

Where, Q_{A_i} is the number of samples correctly predicted as category A_i, and H_{A_i} is the number of total samples correctly predicted as category A_i. The accuracy and F1-score range from 0 to 1. Larger values mean that the model is better at classification.

The doctor recommendation model evaluates the degree of compliance of the recommended doctors by adjusting the influence of each factor. In this paper, a doctor recommendation model is adopted to calculate the real demands of various target users in this paper. The Normalized Discounted Cumulative Gain (NDCG) is employed to indicate the match of the recommended doctor based on the list of doctors to be selected. The higher the value is, the better the recommendation match. The NDCG [26] is calculated as follows:

$$NDCG = \frac{DCG}{IDCG} \tag{15}$$

Where, DCG stands for Discounted Cumulative Gain and IDCG is the maximum DCG value under ideal conditions. The DCG takes into account the influence of the location factor on the recommended results. The order of doctors on the list to be chosen is critical, as different positions will have different contributions. Doctors at the top of the list, on average, have a greater impact than those at the bottom. The DCG is calculated as follows:

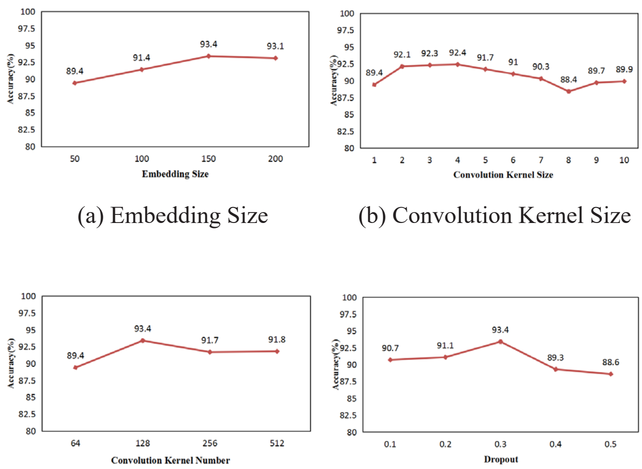
$$DCG_k = \sum_{i=1}^{k} \frac{rel(i)}{\log_2^{(i+1)}} \tag{16}$$

Where, rel(i) is the relevance of the doctor at the position i, and $\log_2^{(i+1)}$ is a discount value.

3.3 Parameter Analysis and Parameter Setting

3.3.1 Parameter Analysis

The selection of hyper-parameters has a significant impact on the model's classification accuracy. In order to further improve the performance of the department classification model, four hyper-parameters Embedding size, Convolution kernel size, Convolution kernel number and Dropout are selected for experimental analysis in this paper.



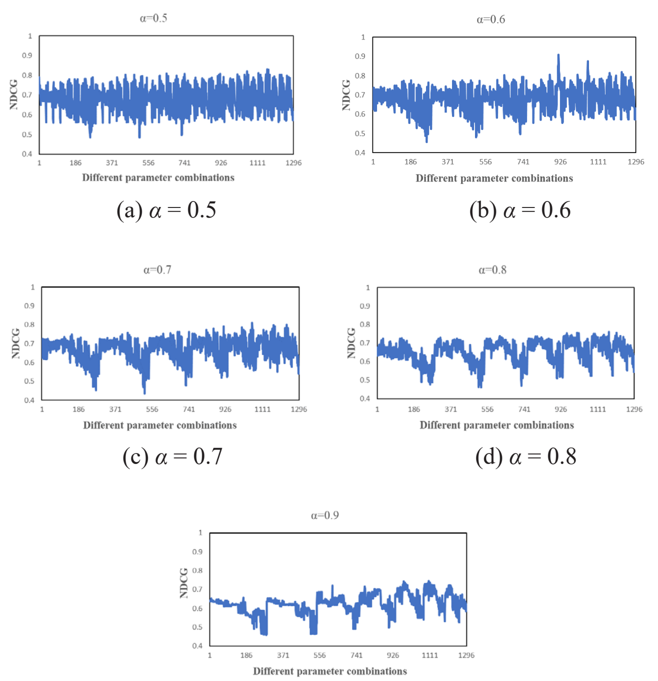
(a) Embedding Size (b) Convolution Kernel Size

(c) Convolution Kernel Number (d) Dropout

Figure 4. Influence of Embedding Size, Convolution Kernel Size, Convolution Kernel Number, and Dropout on accuracy

Figure 4(a) displays that the model is the optimal when the Embedding Size is set to 150. Figure 4(b) illustrates that the model is optimal when the Convolution Kernel Size is set to 4, and that the classification results are similar when the values are 3 and 5. When the optimal convolution kernel size is selected, the combination with convolution kernels with similar classification effect can improve the text classification effect [30]. Therefore, in this paper, the Convolution Kernel Size is set to (3,4,5). When the Convolution Kernel Number is set to 128, Figure 4(c) displays that the model is the optimal. The CNN's convolution kernels mainly extract local features of the text. Since the local features of the patient's chief complaints text are relatively straightforward, the number of convolution kernels required is set to 128. Figure 4(d) reveals that the model has the highest classification accuracy when Dropout is set to 0.3

It is necessary to validate the accuracy of the doctor recommendation model by setting different influencing factors. 80 target users were selected as subjects for the experiment, and information such as the target users' age, gender, and real demands were recorded. Doctor information and patient information were obtained from the information pool. The information was fed into the doctor recommendation models to calculate the top 5 recommended doctors, which was then evaluated by NDCG to ensure the availability and accuracy of the doctor recommendation model. It was concluded from relevant studies [29] and consultations with medical professionals that patient similarity had a greater impact on doctor recommendation matching. Therefore, the patient similarity weights were set to 0.5, 0.6, 0.7, 0.8, and 0.9 in the experiment. The patient preference weights were set to 0.5, 0.4, 0.3, 0.2, and 0.1, respectively. The NDCG with different patient similarity weights are shown in Figure 5(a) to Figure 5(e). When the horizontal coordinates are the different combinations of parameters, this paper does not consider the case where the influence factor is zero. As a result, there are a total of 1296 different combinations of α_1, α_2, α_3, β_1, β_2, and β_3.



(a) $\alpha = 0.5$

(b) $\alpha = 0.6$

(c) $\alpha = 0.7$

(d) $\alpha = 0.8$

(e) $\alpha = 0.9$

Figure 5 Doctor recommendation's NDCG when $\alpha = 0.5$, $\alpha = 0.6$, $\alpha = 0.7$, $\alpha = 0.8$, and $\alpha = 0.9$

As shown in Figure 5(a) to Figure 5(e), when the patient similarity weight was set to 0.6, the NDCG value reached a maximum of 90.7% and the corresponding optimal parameters were as follows: $\alpha = 0.6$, $\beta = 0.4$, $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.2$, $\beta_1 = 0.3$, $\beta_2 = 0.3$, and $\beta_3 = 0.4$. When the patient similarity weight was set to 0.5, the highest value of NDCG reached 83.2%. With a similarity weight of 0.7, the NDCG value reached a maximum of 81.3%; with a similarity weight of 0.8, the NDCG value reached a maximum of 76.1%; with a similarity weight of 0.9, the NDCG value reached a maximum of 74.4%. According to an analysis on the experimental results, patient similarity had a relatively large effect on the doctor recommendation model, among which patient preference had the second largest effect. In patient similarity, the gender factor and the age factor had a similar effect on doctor recommendation matching. Patients of the same gender, on the other hand, had a higher chance of developing a similar disease. Among the factors influencing patient preference, the working years, title, and score of the doctor had a more even effect on doctor recommendation model. Frequently, patients seek out a patient doctor to explain their condition, and they use their doctor's score to figure it out. When selecting a doctor, patients also consider the doctor's title based on their own conditions and the hospital's resources. In addition, the doctor's working years are one of the factors that patients consider when seeking treatment; the longer a doctor has worked, the more experience they tend to have. Therefore, patients will consider a combination of the following three factors: doctor's working years, title, and score. A matched doctor is the most appropriate when a more even weighting is assigned to these three factors.

3.3.2 Parameter Setting

The model's effects are likely to be influenced by a variety of parameters. The model can achieve better results by setting the right parameters. The main parameters for the departmental classification model settings are revealed in Table 2, and the main parameters for the doctor recommendation model settings are displayed in Table 3.

Table 2. Department classification model parameters

Embedding size	Convolution kernel size	Convolution kernel number	Activation function	Learning rate	Dropout
150	(3,4,5)	128	Relu	0.001	0.3

Table 3. Doctor recommendation model parameters

α_1	α_2	α_3	β_1	β_2	β_3	α	β
0.5	0.3	0.2	0.3	0.3	0.4	0.6	0.4

3.4 Experiment Analysis and Discussion of Two Models

3.4.1 Experiment Analysis on Department Recommendation Model

Considering that the text of the patients' chief complaint contains both long and short texts, this paper utilizes CNN [10], CNN_Attention [12], RoBERTa [15], CNN+BiLSTM [16], CNN+RoBERTa [19], and CNN_Attention+RoBERTa for experiments respectively. CNN+BiLSTM indicates text feature extraction using CNN and BiLSTM respectively, followed by feature fusion. CNN+RoBERTa indicates the vectorized representation of the word using the pre-trained model RoBERTa, which is prior to extract features using CNN. CNN_Attention+RoBERTa is the model developed in this paper, which represents the introduction of an attention mechanism to focus on important features after CNN extracts text features, and then fuses them with the features extracted by RoBERTa.

Table 4. Experimental results

Model	Accuracy (%)	F1-score (%)
CNN	84.9	85.1
CNN_Attention	92	92.1
RoBERTa	85.9	86
CNN+BiLSTM	90.5	90.4
CNN+RoBERTa	91.2	91.2
CNN_Attention+RoBERTa	93.4	93.5

The experimental results of each model are revealed in Table 4. The accuracy and F1-score of CNN were 84.9% and 85.1% respectively. Weights were assigned to the text features extracted by CNN as a result of the introduction of the attention mechanism. Focusing attention on the important features of the patient description can better capture the relationship between the contexts in the medical guide text, making it possible to better match the patient description of symptoms with the corresponding consultation department. Therefore, CNN_Attention possesses a better classification effect than CNN. In comparison with CNN, its accuracy and F1-score increased by 7.1% and 7%, respectively. CNN extracts the features mainly through filter windows, focusing on local features of the text, which are not sensitive to text order. To obtain global features, we need to stack them in multiple layers. In this case, the RoBERTa model can be employed to solve this problem. To begin with, RoBERTa is a pre-trained model that can be trained to produce a more closely-related dynamic semantic representation vector. Secondly, RoBERTa is capable to extract full-text contextual features, which can effectively extract long-distance text features. Therefore, RoBERTa classification is better than CNN models. In contrast to CNN, its accuracy and F1-score were increased by 1% and 0.9%, respectively. However, due to the fact that the texts of some patients' chief complaints were not closely linked to each other, the lifting effect was not significant. The CNN+BiLSTM model combines the advantages of CNN and BiLSTM, which can effectively extract textual contextual information. As a result, it outperforms CNN in classification, with the accuracy and F1-score improved by 5.6% and 5.3% respectively. However, The BiLSTM model is unable to learn the connections between words or to reflect the location information of text features. CNN+RoBERTa combines the advantages of CNN and RoBERTa for extracting text features. This method allows for the extraction of additional medical text features. Therefore, it provides better classification results than either the CNN or RoBERTa models alone. In comparison with the CNN model, it increased the accuracy and F1-score by 6.3% and 6.1%, respectively. In contrastto the RoBERTa model, it increased the accuracy and F1-score by 5.3% and 5.2%, respectively. Since CNN is weak at extracting global features of texts, which cannot reflect the importance of text features. As a result, CNN is followed by the introduction of an attention mechanism to focus attention on key features. The features with larger attention values are then fused with the features extracted by RoBERTa, and the fused features are more important for department classification. In comparison with CNN+RoBERTa model, CNN_Attention+RoBERTa increased the accuracy and F1-score by 2.2% and 2.3%, respectively.

3.4.2 Experiment Analysis of Doctor Recommendation Mode

In practical application, the doctor recommendation model should take the real factors into account, such as hospital medical resources, patient needs, and patient satisfaction. Different influencing factors have different impacts on the model. When we select one or several of these factors for model design, it is necessary to analyze the model effect under different influencing factors and different weights. Therefore, in order to validate the accuracy of the proposed doctor recommendation model, three models were utilized in an equivalent experimental setting as follows:

(1) Model A [25]: Word2vec is utilized to preprocess the patient consultation text and doctor history consultation text, and the text words are converted into vectors. Meanwhile, TF-IDF is used to extract key feature words from the doctor history consultation text, and the cosine similarity between the patient and the doctor's history patient is calculated. Consequently, the doctor's recommendation is achieved.

(2) Model B [26]: In combination with deep learning and probabilistic matrix decomposition, CNN is applied to extract features for review information and doctor information. Moreover, Stacked Denoising Autoencoders (SDAE) is utilized to implement pre-trained data representation of the hidden layer to obtain the best initial values of the feature vectors, and the matrix decomposition technology is employed to implement the doctor's recommendation.

(3) Model C [29]: Doctor recommendations are based on factors such as doctor experience, education, ratings, and communication skills. The weight vector based on ranking algorithm assigns different weights according to the importance of the features. The skill score of each doctor is calculated by Multiple Linear Regression (MLR). Recommendation of doctors to patients using KNN algorithm.

Table 5. Comparison of experimental results

Model	NDCG (%)
Model A	83.6
Model B	84.7
Model C	86.1
The proposed model	90.7

The comparison of the proposed model and the other three models is revealed in Table 5. The proposed doctor recommendation model obtained the highest NDCG values, whose NDCG value increased by 7.1% compared to Model A, 6% compared to Model B, and 4.6% compared to

Model C. The Model A mainly focuses on the similarity of consultation text between patients. Model B mainly focuses on the review information and doctor information. Both methods assume that the higher the level of similarity between the conditions is, the higher the matching rate will be. The impact on medical resources and patient preferences in practical applications is ignored. Model C not only focuses on patient text information, but also introduces plentiful features (such as doctor's experience, education, and rating), and assigns different weights according to the importance of the features. The factors considered in this way are more comprehensive, so that the NDCG value of Model C is higher in contrast to Models A and B, which is more in line with the demands in practical applications. However, when there are more influencing factors, we must consider how to select the key factors among them. For instance, in addition to the similarity of counselling texts between patients, gender and age between patients are also influential factors for condition matching. In addition, the actual situation of medical resources is not considered in Model C. If patients select doctors with higher titles, it is likely to lead to the tension of medical resources. According to the requirements of practical application, the model proposed in this paper assigns reasonable weights to various influencing factors and achieves better results in the experiment.

3.4.3 Discussion

The results reveal that the models proposed in this paper are effective. For the department classification model, the prediction performance depends largely on the model's understanding of the patients' chief main complaint texts. Therefore, when the model is designed, it needs to fully consider the characteristics of the length of the patients' main complaint texts. The CNN model proposed by Kim [10] achieved an 85.1% F1-score. CNN has a good command of extracting local features of text. However, CNN captures textual information mainly through convolutional operations, which is prone to the loss of key information and location information. Liu et al. [15] achieved an 86% F1-score adopting the RoBERTa model, a pre-trained model based on a large-scale corpus that understands textual semantics and contextual information. However, pre-training data can lead to performance degradation if they do not match due to the target task. Alshubaily et al. [12] achieved a 92.1% F1-score applying the CNN_Attention model and achieved better classification results. This model introduces an attention mechanism after CNN to assign different features to different features and focus on text critical features. This suggests that the dataset employed in this paper is likely to have a higher number of short texts. Therefore, using this model could achieve better results. However, due to the fact that some long text still exists in the dataset, using this model can result in partial loss of information. Liu et al. [16] achieved a 90.4% F1-score using the CNN+BiLSTM model. CNN+BiLSTM combines the advantages of CNN in capturing local features and BiLSTM in capturing long-range sequential features. However, BiLSTM can result in partial loss of information due to the inability to fully utilize the information of the whole text when processing long text. Mu et al. [19] achieved a 91.2% F1-score applying the CNN+RoBERTa model. CNN+RoBERTa introduces a pre-training model that achieves a rich semantic representation of words, which can effectively improve the performance of text classification models. However, when the text features are plentiful and complex, it is not difficult to lead to insufficient attention to the key features of the text. The proposed model combines the richness of Word2vec and RoBERTa for word vector representation, and integrates the advantages of CNN_Attention and RoBERTa for extracting short and long texts. As a result, the proposed model obtains a 93.5% F1-score.

The patient needs to be matched with the suitable doctor after locating the doctor in the department through the department recommendation model. Case similarity between patients, doctor's professionalism, and patient's preference are the influencing factors of the doctor recommendation model. When the effect of one or more of these factors is ignored, it may influence the accuracy of the doctor's recommendation. Kou [25] and Yan et al. [26] match doctors based on the similarity of conditions between patients, which is meaningful because of the following causes: in practical application, in view of the patient's condition, the patient cannot select a doctor arbitrarily, otherwise the patient's condition will be delayed. In addition, from the consideration of medical resources, the patient can't blindly choose a doctor with a high title, which will lead to the tension of the doctor's resources. Singh et al. [29] designed a model incorporating characteristics, such as doctor experience, doctor ratings, and communication skills. The NDCG value obtained by this method was 86.1%. This approach takes a variety of influencing factors into account, and allows for more objective matching of doctors. However, from the consideration of the patient's preference, the patient can't select the doctor against his will, which will lead to a decrease in the patients' satisfaction. It can be concluded from the above content that the accuracy of doctor's recommendation is influenced by a variety of factors. Therefore, the proposed model integrates both patient similarity and patient preference, and assigns different weights according to the importance of features to achieve accurate doctor-patient matching. The model proposed in this paper achieved the NDCG value of 90.7%. In the actual medical consultation service, the model can meet the demands of matching patients' conditions and doctors' profession, alleviate the pressure on medical resources, and improve patients' satisfaction.

Our results demonstrate good performance metrics, while they are not perfect. The department classification model has only been validated in the classification of internal medicine departments in hospitals, where different diseases may have the same symptom descriptions. However, this paper does not validate the effectiveness of feature extraction in this regard. In addition, this paper only discusses the impact of six factors on the doctor recommendation model. However, doctor recommendation may also be related to more secondary factors, such as doctors' communication skills and staff behavior. In this way, the issue of assigning weights to other minor factors and the impact on the model needs to be further explored.

4 Conclusion and Future Work

This paper proposes the utilization of a department classification model to locate doctors, and applies a doctor recommendation model combining patient similarity and patient preference so as to produce accurate doctor recommendations. With an accuracy of 93.4% and F1-score of 93.5%, the proposed department classification model is more effective in classifying the dataset. This paper selects a number of target users for experiments to validate the practicality and effectiveness of the doctor recommendation model. The NDCG of doctor's recommendation reached 90.7%.

Further validation of the model's validity for other departmental classifications will be required in the future. Meanwhile, in order to design a doctor recommendation model that is more responsive to real-world demands, further investigation is still needed on the impact of other factors on doctor recommendation.

Acknowledgements

This research was funded by Natural Science Foundation of Jiangsu Higher Education Institutions of China (Grant 23KJD520011), Research topic on educational informatization in Jiangsu Higher Education Institutions (Grant 2023JSETKT126), and Nantong Institute of Technology Research Project (Grant 2022XK(J)15).

References

[1] X. Bu, L. Lu, Z. Zhang, Q. Zhang, Y. Zhu, A general outpatient triage system based on dynamic uncertain causality graph, *IEEE Access*, Vol. 8, pp. 93249-93263, May, 2020.

[2] P. Guo, W. Deng, Design and implementation of intelligent medical customer service robot based on deep learning, *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, Chengdu, China, 2019, pp. 37-40.

[3] M.-K. Dalal, M.-A. Zaveri, Automatic text classification: a technical review, *International Journal of Computer Applications*, Vol. 28, No. 2, pp. 37-40, August, 2011.

[4] H. Xu, M. Jiang, M. Oetjens, E.-A. Bowton, A.-H. Ramirez, J.-M. Jeff, M.-A. Basford, J.-M. Pulley, J.-D. Cowan, X. Wang, M.-D. Ritchie, D.-R. Masys, D.-M. Roden, D.-C Crawford, J.-C. Denny, Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin, *Journal of the American Medical Informatics Association*, Vol. 18, No. 4, pp. 387-391, July, 2011.

[5] T. Botsis, M.-D. Nguyen, E.-J. Woo, M. Markatou, R. Ball, Text mining for the vaccine adverse event reporting system: medical text classification using informative feature selection, *Journal of the American Medical Informatics Association*, Vol. 18, No. 5, pp. 631-638, September, 2011.

[6] J. Krämer, J. Schreyögg, R. Busse, Classification of hospital admissions into emergency and elective care: a machine learning approach, *Health care management science*, Vol. 22, No. 1, pp. 85-105, March, 2019.

[7] C.-C. Aggarwal, C.-X. Zhai, A survey of text classification algorithms, *Mining text data*, 2012, pp. 163-222.

[8] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *nature*, Vol. 521, No. 7553, pp. 436-444, May, 2015.

[9] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, Deep learning--based text classification: a comprehensive review, *ACM computing surveys (CSUR)*, Vol. 54, No. 3, pp. 1-40, April, 2022.

[10] Y. Kim, *Convolutional neural networks for sentence classification*, arXiv preprint arXiv:1408.5882, September, 2014. https://arxiv.org/abs/1408.5882

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.-N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Proceedings of the 31st International Conference on Neural Information Processing System*, Long Beach, CA, USA, 2017, pp. 6000-6010.

[12] I. Alshubaily, *Textcnn with attention for text classification*, arXiv preprint arXiv:2108.01921, August, 2021. https://arxiv.org/abs/2108.01921

[13] L. Xiao, G. Wang, Y. Zuo, Research on patent text classification based on word2vec and LSTM, *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, 2018, pp. 71-74.

[14] G. Xu, Y. Meng, X. Qiu, Z. Yu, X. Wu, Sentiment analysis of comment texts based on BiLSTM, *IEEE Access*, Vol. 7, pp. 51522-51532, April, 2019.

[15] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692, July, 2019. https://arxiv.org/abs/1907.11692

[16] Z. Liu, D.-G. Zhang, G.-Z. Luo, M. Lian, B. Liu, A new method of emotional analysis based on CNN–BiLSTM hybrid neural network, *Cluster Computing*, Vol. 23, No. 4, pp. 2901-2913, December, 2020.

[17] M. Wang, Q. Cai, L. Wang, J. Li, X. Wang, Chinese news text classification based on attention-based CNN-BiLSTM, *PPR 2019: Pattern Recognition and Computer Vision*, Wuhan, China, 2020, pp. 110-117.

[18] J. Deng, L. Cheng, Z. Wang, Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification, *Computer Speech & Language*, Vol. 68, Art. No. 101182, July, 2021.

[19] Z. Mu, S. Zheng,Q. Wang, ACL-RoBERTa-CNN text classification model combined with contrastive learning, *2021 International Conference on Big Data Engineering and Education (BDEE)*, Guiyang, China, 2021, pp. 193-197.

[20] X. Chen, P. Cong, S. Lv, A long-text classification method of Chinese news based on BERT and CNN, *IEEE Access*, Vol. 10, pp. 34046-34057, March, 2022.

[21] L. Yong, J. Huang, A search of personalized doctor recommendation algorithm, *Information &*

Communications, Vol. 2, No. 13, pp. 181-188, April, 2018.

[22] Y.-F. Huang, P. Liu, Q. Pan, J.-S. Lin, A doctor recommendation algorithm based on doctor performances and patient preferences, *2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)*, Chengdu, China, 2012, pp. 92-95.

[23] M. Waqar, N. Majeed, H. Dawood, A. Daud, N.-R. Aljohani, An adaptive doctor-recommender system, *Behaviour & Information Technology*, Vol. 38, No. 9, pp. 959-973, June, 2019.

[24] F. Narducci, C. Musto, M. Polignano, M.-d. Gemmis, P. Lops, G. Semeraro, A recommender system for connecting patients to the right doctors in the healthnet social network, *Proceedings of the 24th international conference on World Wide Web*, Florence, Italy, 2015, pp. 81-82.

[25] X. Kou, Personalized doctor recommendation method based on patient online consultation text, *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Fuzhou, China, 2021, pp. 31-34.

[26] Y. Yan, G. Yu, X. Yan, Online doctor recommendation with convolutional neural network and sparse inputs, *Computational Intelligence and Neuroscience*, Vol. 2020, Article No. 8826557, October, 2020.

[27] Z. Zheng, Z. Qiu, H. Xiong, X. Wu, T. Xu, E. Chen, X. Zhao, DDR: Dialogue based doctor recommendation for online medical service, *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington, USA, 2022, pp. 4592-4600.

[28] Y. Yang, J. Hu, Y. Liu, X, Chen, Doctor recommendation based on an intuitionistic normal cloud model considering patient preferences, *Cognitive computation*, Vol. 12, No. 2, pp. 460-478, March, 2020.

[29] H. Singh, M.-B. Singh, R. Sharma, J. Gat, A.-K. Agrawal, A. Pratap, Optimized doctor recommendation system using supervised machine learning, *Proceedings of the 24th International Conference on Distributed Computing and Networking*, Kharagpur, India, 2023, pp. 360-365.

[30] X. Li, M. Cui, J. Li, R. Bai, Z. Lu, U. Aickelin, A hybrid medical text classification framework: Integrating attentive rule construction and neural network, *Neurocomputing*, Vol. 443, pp. 345-355, July, 2021.

Biographies

Hao Chen received his ME degree from the School of Energy and Power Engineering, Yangzhou University, Yangzhou, China in 2016. He is currently a lecturer at Nantong Institute of Technology. His research interests include computing systems and artificial intelligence.

Wenye Gu received her bachelor's degree from the medical school, Nantong University, Jiangsu, China in 2014. She is currently a nurse-in-charge of the Oncology Radiotherapy Department at Affiliated Hospital of Nantong University.

Haifei Zhang received his ME degree from the School of Computer Science and Technology, Soochow University, Suzhou, China in 2005. He is currently a Professor at the Nantong Institute of Technology. His research interests include deep reinforcement learning and GIS.

Qiong Zhang is currently a PhD student in Fusion IT at Mokwon University, Daejeon, Korea. He is currently a lecturer at Nantong Institute of Technology. His research interests include computing vision and digital image processing.

Xiaowei Ding is currently a senior engineer at Jiangsu Chaoshu Information Technology Co., Ltd.