Chinese Patent Medicine Recommendation Algorithm Based on DPCNN-DeepFM

Yunpeng Huang, Huihua Yang, Weiguang Wang, Feng Li, Xiaofeng Liu and Jiahui Ouyang

October 12, 2023

Chinese Patent Medicine recommendation algorithm based on DPCNN-DeepFM

Yunpeng Huang[a], Huihua Yang*[a], Weiguang Wang[b], Feng Li[b], Xiaofeng Liu[b], Jiahui OuYang[c]
[a] School of artificial intelligence,Beijing University of Posts and Telecommunications, Beijing, China, 100876; [b]College of traditional Chinese Medicine,Beijing University of Chinese Medicine, Beijing, China, 100029; [c]College of Graduate School Medicine,Beijing University of Chinese Medicine, Beijing, China, 100029

ABSTRACT

In this paper, we combine the embedded deep pyramid network and the deep FM recommendation model to complete the recommendation of proprietary Chinese medicines. Using real electronic medical records provided by Beijing University of Chinese Medicine and the Institute of Computer Science, Chinese Academy of Sciences as data sources; according to the patient's main complaint, symptoms, tongue, pulse, age, gender, and other characteristics; using a deep pyramid network based on embedded patient feature fusion to predict the patient's card Candidate categories, realize long-distance text association through text region embedding and fixed number of feature maps; and use deep FM model to process the sparse and dense features of the patient with the FM layer and the DNN layer separately, and finally real-time features and predictions for the patient The syndrome types are sorted. After comparative experiments and training, the AUC index is 78.90%, and the ACC value is 83.26%. Its overall performance is better than other RNN, CNN, and fast text network models. The experimental results show that the recommendation algorithm of Chinese patent medicine based on DPCNN-DeepFM has certain reference significance for the auxiliary clinical diagnosis of patients.

Keywords: TCM diagnosis, Chinese Patent Medicine, recommendation system, DPCNN, DeepFM

1. INTRODUCTION

Since artificial intelligence has become a hot spot in recent years, many studies have proposed related algorithm models in the field of drug recommendation. Beckert [1] uses a demographic-based collaborative filtering algorithm to evaluate clinical interview samples of patients with autoimmune skin disease psoriasis. The data of this algorithm is mainly derived from the health records provided by the Dermatology Clinic and Polyclinic of the Dresden University Hospital. It is mainly composed of consultations of 213 patients with various types of psoriasis and stored in the local MySQL database. Use a collaborative filtering algorithm to realize drug-related recommendations. Balvert [2] uses a combination of support vector machine nuclear regression and random forest to test the effects of 43 drugs on 1018 cancer cell lines and obtain the best combination of drug prescriptions in cancer treatment based on genomic information. Recommend system. Zhou Xiling [3] cited the similarity of user demographic attributes based on calculating the similarity of user ratings.

The weighted linear fusion of demographic attribute similarity and user score similarity is performed to obtain user similarity. Finally, the similarity set is selected by the user similarity to generate the recommendation result of the target user. In drug recommendation algorithms, deep learning methods have more powerful characterization capabilities than traditional methods, and deep features can be extracted through nonlinear networks. As a patient and related drug data are in the field of medicine, there are few data publicly available on the Internet and a large number of professionals are required to mark them. Usually, patients need to follow up for a long time to score drugs, and the collection period is long, which may easily cause data loss.

In the case of a small amount of data, the effect of the depth model is often poor. Based on the analysis and research of related issues, this article adopts the method of text classification to deal with drug recommendations, based on the deep pyramid text classification convolutional network model, and refers to the diagnosis and treatment guidelines of Chinese medicine and the diagnosis and treatment of common diseases in Chinese medicine. tongue, pulse, symptoms, age, gender, and other characteristics. Select specific 7 types of TCM internal medicine diseases and 43 types of syndromes

*yhh@bupt.edu.cn; phone 86 13718680586; https://teacher.bupt.edu.cn/yanghuihua

for processing and analysis, build a TCM syndrome classification model and use the prediction results to complete the recommendation of proprietary Chinese medicines.

1.1 Bag of words model

In natural language processing, the traditional method calculates the frequency of related words by counting the number of occurrences of a word, thereby expressing the information contained in the word. In a common language model, for a text sequence $S = w_1, w_2, \ldots, w_T$, its probability can be expressed as:

$$P(S) = \prod_{t=1}^{T} p\,(w_t | w_1, w_2, \ldots, w_{t-1}) \tag{1}$$

Because the amount of information contained in statistical features is too small. This severely limits the rich semantic features of words and their representation in depth. With the research of many scientific researchers on feature representation. The word vector model has been proven to be used in feature representation. There are two common representation methods: 1.one-hot vector representation; 2. distributed representation. The one-hot vector is very common in natural language processing because its advantage is that it is simple and fast. By counting the number n of all different words, construct a vector where the j-th position is 1 and the rest are all 0 to represent each word In natural language processing, the traditional method calculates the frequency of related words by counting the number of occurrences of a word, thereby expressing the information contained in the word.

But the one-hot vector has its limitations: the vocabulary in some typical fields contains as many as tens of thousands of words, and each word needs to be encoded. Therefore, these vectors are usually large in length and are very easy to cause dimensional disasters, and also ignore important word-order relations in the text. Take the three texts of " no headache ", " mild headache " and " severe headache " as examples. The size of the dictionary constructed by counting the words is 4, where "no" can be represented by the vector $[1,0,0,0]$, and " headache" can be represented by the vector $[0,1,0,0]$ to represent. The three paragraphs of text are:

Table 1. One Hot Text Vector Representation

Original text	**Text vector representation**
no headache	$[1,1,0,0]$
mild headache	$[0,1,1,0]$
severe headache	$[0,1,0,1]$

When constructing two paragraphs of text, the model hopes that the vector representations of " no headache " and " mild headache " are closer to each other. But when using one-hot vector representation, the generated word vector is two independent vectors and has no semantic relationship, which brings many problems to model training and subsequent downstream tasks.

1.2 Distributed representation model

Distributed representation is a low-dimensional dense word vector obtained through training, Google proposed Word2vec [4] in 2013. Word2vec mainly has two methods for training word vectors, the Continuous Bag-of-Words Model (CBOW) and the Continuous Skip-gram Model, Skip-Gram). The CBOW model reads the words in the context window and tries to predict the most likely center word.

The Skip-Gram model is just the opposite, predicting the context word based on the given center word. Taking CBOW as an example, as shown in the figure above, the model predicts the target word by trying to understand the context of surrounding words. Given the $\mathrm{Context}(w)$ of word w, word w in the model is a positive sample, and the rest of the words are negative samples. Given a negative sample subset $\mathrm{neg}(w) \neq \varnothing$, By maximizing the objective function:

$$g(w) = \mathop{\Pi}_{u \in neg(w)} p\big(u | Context(w)\big) \tag{2}$$

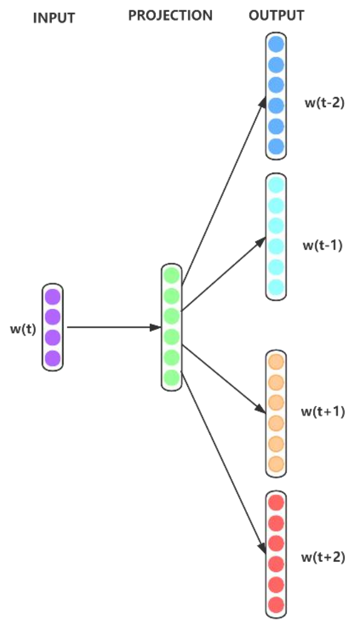

Figure 1. Continuous skip word model

Then the word vector of the word can be obtained. At the same time of calculation, it is also necessary to average the word vectors of multiple context words to obtain the same fixed-length vector as the hidden layer. The objective function of a given word sequence $[w_1,w_2,w_3,\cdots,w_T]$, CBOW is the negative log-likelihood function of a set of context words, and the probability of generating the central word is as follows:

$$p(w_O|w_i) = \frac{\exp\left(v_{w_O}^{'\,T} v_{w_i}\right)}{\sum_{w=1}^{W} \exp\left(v_w^{'\,T} v_{w_i}\right)} \tag{3}$$

Among them, $v_w^{'}$ and v_{w_i} represent the output vector and input vector of word w, respectively. Through the training of a large corpus, the weight parameters between the input layer and the mapping layer are taken out by the backpropagation algorithm to obtain the trained word vector.

2. EMBEDDED FEATURE FUSION MODEL BASED ON DPCNN-DEEPFM

2.1 Summary

This paper proposed a method based on DPCNN [5] to solve the problem of Chinese patent medicine recommendation. The model integrates the patient's embedded characteristics to complete the TCM syndrome classification algorithm. According to the predicted syndrome results, it completes the recommendation of Chinese patent medicine.

The model can be divided into three parts: the first part is the text region embedding, the symptom part word vector is input, and the text convolution block is used to extract high-order features; the second part is the short-path connection, which mainly completes the accuracy of the model by deepening the number of layers The improvement and realization of long-distance patient feature correlation; the third part is repeated down-sampling, through repeated down-sampling operations, the parameters of the long-distance feature sequence are simplified, and the calculation efficiency of the model is improved.

2.2 Text area Embedding

In the model in this article, since the electronic medical record of Chinese medicine contains a large amount of patient symptom information, when the patient's symptom features are represented by the Word2Vec algorithm in distributed vectorization, it does not directly use a single word vector for feature extraction through convolutional blocks. Instead, the text area embedding layer is transformed into a text area vector of multiple words to extract patient features. This can effectively include more semantic information in the input part, which makes the features extracted by the model more complex.

In the calculation process of the model, the number of filters and the embedding vector dimension are pre-set hyperparameters. When the number of filters is N and the dimension of embedding is K, the text convolution area adopts 2d convolution, and the input size is $[1, \mathrm{N}, (3 \times \mathrm{K})]$, where 1 is the number of input channels size. Since it is a text vector, set its value to 1. In the parameters set by the model $(3 \times \mathrm{K})$ is the size of the convolution kernel. The input is calculated by calculating the word vectors around the central word, and the objective optimization function is:

$$\sum_{i,j} a_{i,j}(z_i[j]-p_i[j])^2 \quad (4)$$

In the formula, $z_i[j]$ represents the vector of the word bag model with word j, $a_{i,j}$ is the weight, and $p_i[j]$ is the output of this time.

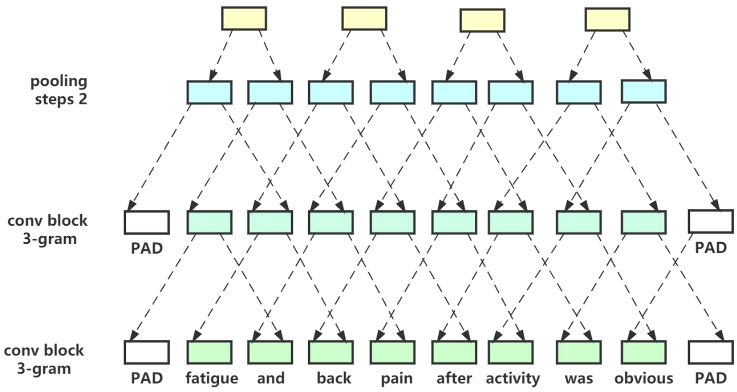

Figure 2. . Schematic diagram of convolutional block pooling

2.3 Short-circuit connection

Since the convolutional neural network was widely used in the image field at first, and experiments have shown that the deep network model will significantly improve the feature extraction effect of the model. However, if the deep model is directly constructed because the initial gradient is too small, the gradient value of each layer model will be very small, making the model training time longer and the effect is not good. Drawing lessons from the residual structure introduced by ResNet, this article uses a short-circuit connection operation, and the output of the short-circuit connection can be expressed as:

$$o_1 = z + f(z) \quad (5)$$

Where f is the jump layer, and z is the output after the text embedding layer. Since the model uses equal-length convolution, there is no need to frequently change the size at the output. By connecting the output of this layer with the directly connected layer, the exchange of gradients between the various layers of the model is accelerated, thereby accelerating the training efficiency of the model.When initializing CNN, the weights of each layer are often initialized to very small values, which leads to the initial network, the input of almost every subsequent layer is close to 0, and the network output at this time is meaningless;The small weight hinders the propagation of the gradient, so that the initial training phase of the network often takes a long time to start;Even if the network is started up, because the affine matrix (the connecting edge between every two layers) in the deep network is approximately continuous, the network is very prone to gradient explosion or dispersion problems during the training process.Using Shorcut connection, this can greatly alleviate the problem of gradient disappearance.

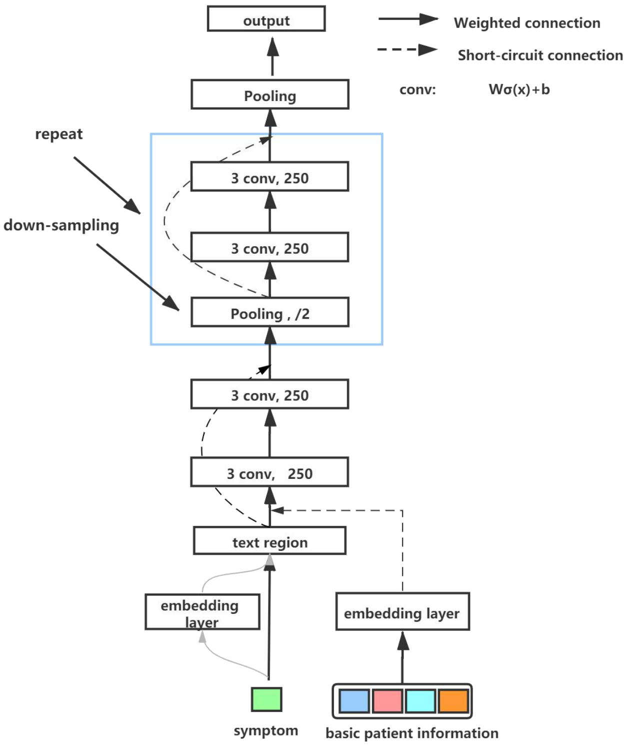

Figure 3. Embedded Feature Fusion Model Based on DPCNN

2.4 Repeated downsampling

The model integrates the symptom-related text information into the text area embedding layer, and other classification information uses the Word2Vec algorithm to obtain the corresponding vector.

The generated corresponding vector is as follows:

$$M_{1:I} = [m_1, m_2, \cdots m_I] \tag{6}$$

Among them, m_I is the first classification feature, and the feature vector input to the convolution block is the feature vector fused with other classification information. The formula is as follows:

$$X' = X + M_{1:I} \tag{7}$$

When the model uses the convolution block for feature extraction, in order not to change the input structure of the corresponding data, considering that the input size does not need to be adjusted after the model, this article uses the equal-length convolution in the convolutional neural network to operate, through the 0-filling operation Complete the filling of the model shape. The characteristics of the output are:

$$\mathrm{W}\sigma(\mathrm{X} + \mathrm{M}_{1:I}) + \mathrm{b} \tag{8}$$

Where W is the weight of the current layer, b is the bias term, and they are all parameters that can be trained. After passing the output of the two convolutional blocks, the length of the feature sequence is reduced layer by layer by repeated down sampling processing, and finally, the pooling layer is used for maximum pooling, which is processed by the nonlinear activation function Relu to obtain the TCM syndrome type The classification results are output. When the model is trained, the optimizer uses Adam to optimize. The loss function used in this model is the cross-entropy loss function, and its formula is expressed as follows:

$$-\frac{1}{n}\sum_x [yln\sigma(z) + (1-y)ln(1-\sigma(z))] \quad (9)$$

3. CHINESE PATENT MEDICINE RECOMMENDATION RANKING MODEL BASED ON DEEPFM

3.1 Summary

Through the deep pyramid recall model of embedded fusion features, we can get the most relevant candidate set of syndrome types. However, using candidate sets and real-time disease information directly often ignores the cross features and other information features of patients. In order to use more detailed combination characteristics of patients, and add the consideration of patient related information into the rapidly screened user ID. In this paper, after building the recall model, we add the ranking model to make the effect of recommendation more accurate. In the candidate syndrome type ranking model, the features are mainly divided into three parts. The first part is the characteristics of patients' symptoms, which mainly includes the characteristics of patients' basic information and syndrome types; The second part is the characteristics of patients, mainly including the main symptom information of patients this time. The basic information of patients includes age, gender, tongue condition, pulse condition and other basic information; The third part is the related situation of syndrome type, mainly for the explanation of syndrome type and the type of medication. After continuous development, the features constructed by candidate ranking model from linear model to depth model become more and more complex, which makes the model can mine more deep level feature information and make the recommendation effect more accurate.

3.2 TCM diagnosis and treatment system based on TCM diagnostics

As the actual process involves a lot of medical knowledge, through consulting professionals and consulting relevant literature, the theoretical basis of this paper is the TCM diagnosis and treatment system proposed by TCM diagnostics, through the combination of disease, syndrome and disease to recommend drugs to patients.

In traditional Chinese medicine, the diagnosis and treatment system of traditional Chinese medicine is mainly composed of the combination of disease, syndrome and disease. Through combing the essence of disease, traditional Chinese medicine mainly consists of syndrome differentiation, syndrome differentiation and disease differentiation. If we want to establish the diagnosis and treatment system of traditional Chinese medicine, we must first understand the meaning and relationship among "disease", "syndrome" and "disease".

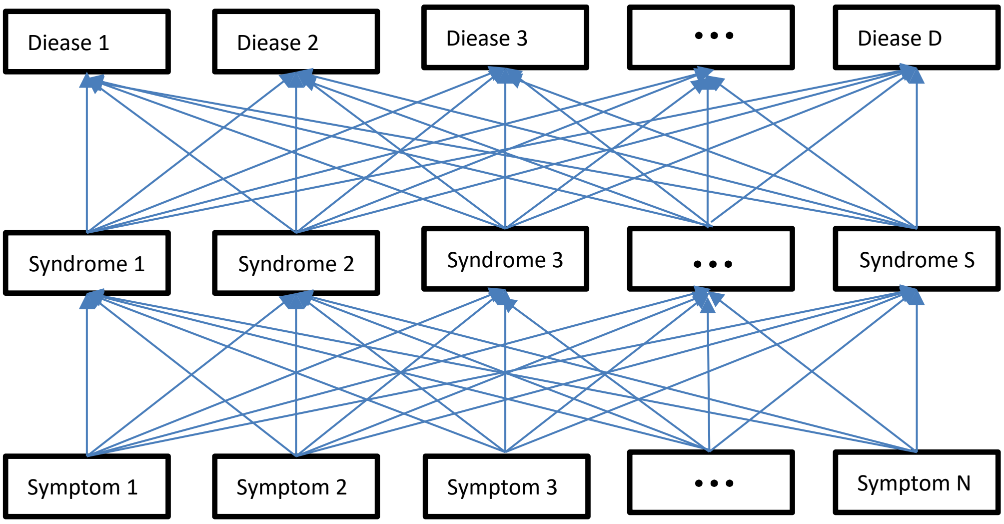

Figure 4. Disease syndrome syndrome model of TCM diagnosis and treatment system

Symptoms, syndromes and diseases are interrelated and complicated in the system of disease syndrome syndrome diagnosis and treatment. A syndrome is often composed of multiple symptoms, and a disease often has multiple syndromes. Each symptom has a very important impact on the judgment of patients' diseases. By analyzing the various symptoms of patients and using the theory of traditional Chinese medicine to infer the relevant syndromes, we can determine the patients' diseases of traditional Chinese medicine and realize the patients' recommended medication.

3.3 DeepFM

In the calculation, the recommendation ranking model of Chinese patent medicines based on DeepFM needs to encode discrete features such as patient gender and syndrome type. The syndrome type of the input port is the recall candidate set derived from the deep pyramid text convolutional network with embedded fusion features. It can be clearly seen from Figure 7 that the real-time features are the current basic information of the patient and the syndrome recall of the previous model. result. By embedding vectorization of real-time features, part of the sparse layer is crossed through the FM layer, mainly to extract low-level feature information in the patient's medical record. The other part passes through the dense layer and uses DNN to extract high-order features. The features of the dense layer are input into the hidden layer.

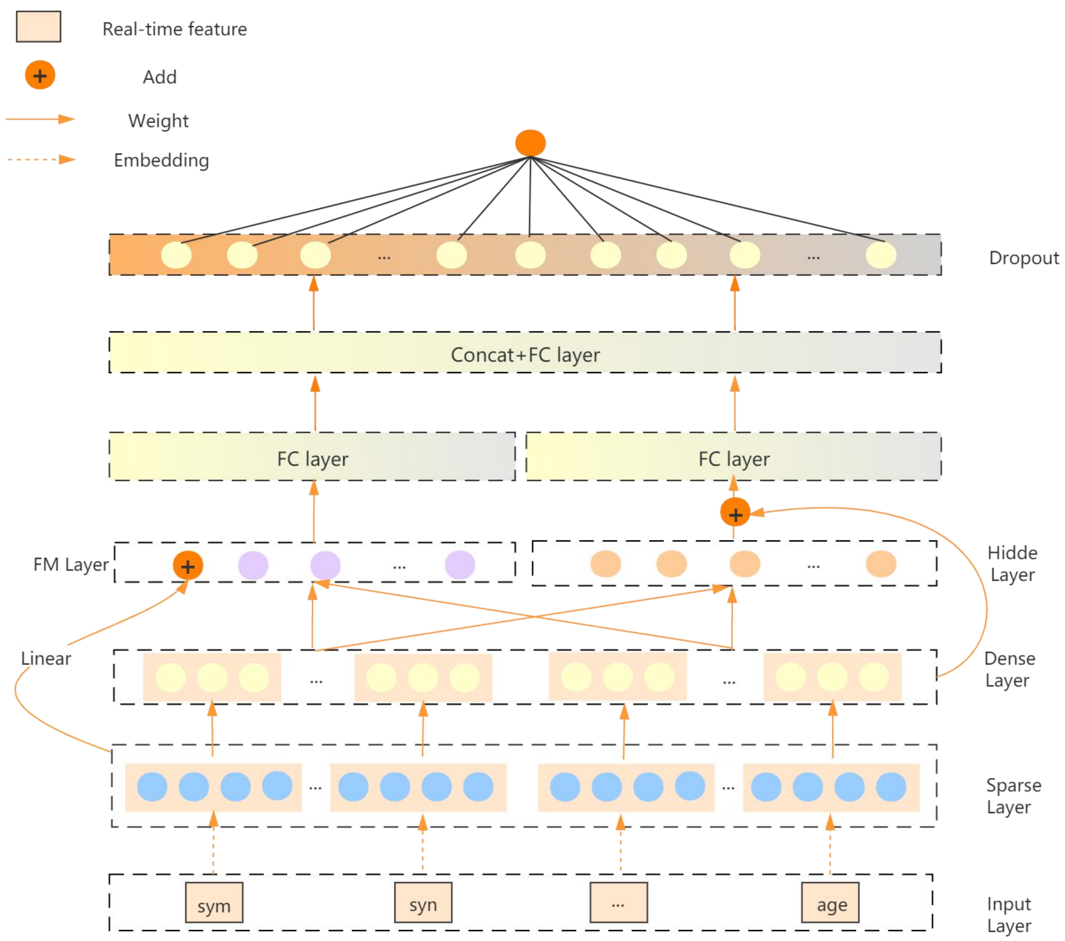

Figure 5. Recommendation model of Chinese patent medicine based on DeepFM

Therefore the recommended ranking model of Chinese patent medicines based on DeepFM used in this article combines the advantages of both. And in the dense layer, a part of the feature vector passes through the hidden layer as input, and the other part is connected with the output of the hidden layer by residual error and then added to the sum. Then use the output of the FM layer and the result of the hidden layer and the residual of the dense layer as input, pass through the two fully connected layers and perform splicing processing. The output vector uses Dropout to randomly remove neurons, and finally outputs the predicted value of the syndrome type through the activation function and sorts the results to complete the overall prediction of the Chinese patent medicine recommendation ranking model based on the DeepFM model.

4. EXPERIMENTAL ANALYSIS

To verify the effect of the model, this model is compared with the other 5 mainstream classification models. The pre-trained word vector loaded in this article is an open-source, large-scale, high-quality Chinese word vector data set published by Tencent AI Lab. The data set contains more than 8 million Chinese words, and the decompressed size is 15.5G. From the comparison of model training loss in Figure 5, the training network fluctuates at about 6000 steps. By checking the training and test data, it is found that the various syndromes of constipation are less than other data, and it is predicting this type of syndrome. As a result, the loss value becomes larger. Compared with other models, the model based on DPCNN has the smallest loss value.

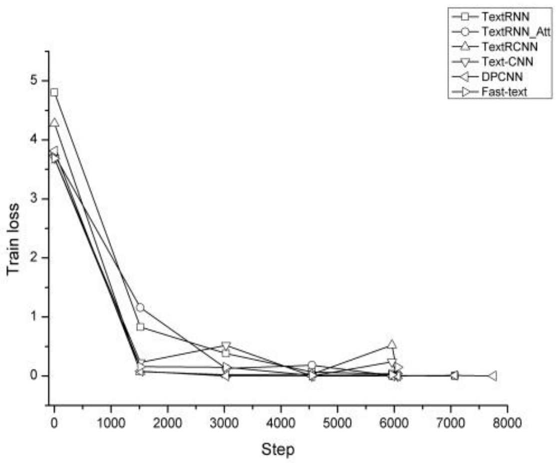

Figure 6. The training loss of model

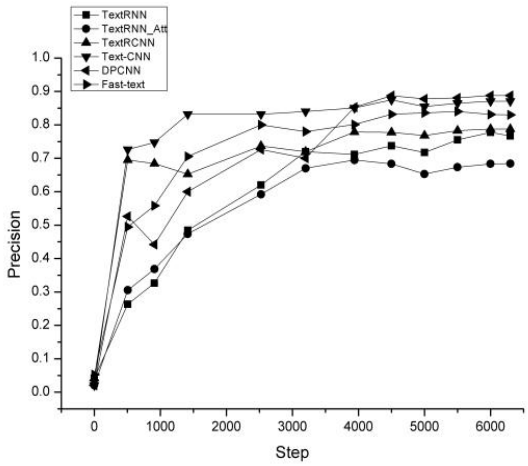

Figure 7. The comparison chart of accuracy rate of TCM syndrome classification

In actual operation and comparing the training time of 6 model tests, TextRCNN takes the shortest time to complete one training, but on the test set, the accuracy of TextRCNN is not high, and the F1-score of DPCNN can be derived from Table 2. best effect. Through analysis, the difference between DPCNN and TextRCNN is that although they have good effects on long-distance dependence, due to the deeper network structure of DPCNN and multi-layer pooling, higher-order complex semantic information can be effectively extracted. But compared to TextRNN, the model with the attention mechanism has lower accuracy. Usually, when the attention mechanism is added, the model indicators generally increase.

Table 2. The evaluation results of Chinese patent drug

Models	Precision	Recall	F1-score
TextRNN	0.7875	0.6444	0.7088
TextRNN_ATT	0.6833	0.5369	0.6013
TextRCNN	0.7873	0.6404	0.7063
Text-CNN	0.8750	0.7414	0.8027
Fast-text	0.8417	0.6891	0.7578
DPCNN	**0.8875**	**0.7517**	**0.814**

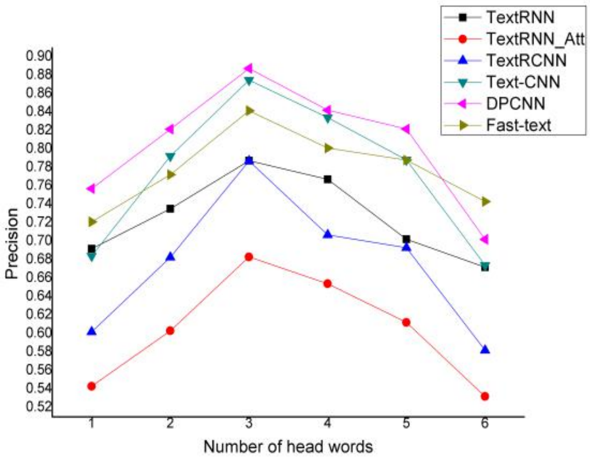

Figure 8. The embedding number of head words

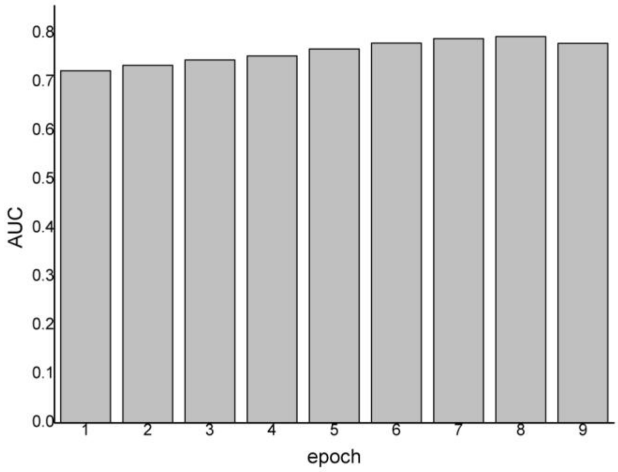

Figure 9. The change of AUC index

In this paper, we test the accuracy of AUC prediction by adjusting the length of embedding layer. When the number of neurons in the hidden layer is 250, we calculate the average AUC value by training 8 epochs of the model. The prediction results are as follows:

Table 3. The relationship between the size of embedding layer and AUC value

Embedding size	AUC(%)	ACC(%)
16	65.31	70.65
32	70.45	76.84
64	73.46	79.63
128	75.93	81.49
256	**78.90**	**83.26**
512	78.03	84.07

When the model embedding layer size is 256, the network sorting model has the best effect, so in the sorting process, the model chooses to use the 256 dimension of embedding dimension for the actual reasoning sorting calculation. When the size of embedding layer is 512, the AUC index of the model begins to decline. The AUC can reach 78.9% and the accuracy can reach 83.26% when the size is 256, so it can be a good recommendation for Chinese medicine.

However, the effect of deep learning models often depends on the structure of the data, and because the model classifier does not learn more complex features when the amount of data is not large, the complexity of the model increases and the effect is not good.

5. CONCLUSION

This paper presents a Chinese patent medicine recommendation algorithm for patients' symptoms. Based on the text classification model DPCNN, it discusses the three links of model-related theoretical analysis, model algorithm design, and experimental verification. The experimental results show that the prediction f1 index of the TCM syndrome type classification of the specific seven diseases by this method model reaches 81.4%. The patient selects the relevant symptoms and basic information, infers and calculates the syndrome type classification, and uses the predicted results to recommend medication.

Although it has certain advantages over mainstream models, the recommended algorithm model of Chinese patent medicines in this article is only for specific diseases. Moreover, due to the uneven distribution of the original data, even after under-sampling and over-sampling, some noise and unbalanced samples will still affect the classification effect; these are the shortcomings of this research and the direction to be improved.

Finally, with the development of the modernization of Chinese medicine, more and more deep learning models are used in the field of Chinese medicine. It is also hoped that the model in this article can provide a certain reference for the research of drug recommendation algorithms.

6. ACKNOWLEDGMENT

This work was supported by 2018YFC1704100, National Key R&D Program of China：Research on Inheritance and Innovation of Experience (Integration of Tao and Shu) of Illustrious Senior Traditional Chinese Medicine Practitioners by adopting Multiple Research Methods. The author thanks Beijing University of Chinese Medicine for providing real medical records, and help this article mark a lot of original data and theoretical guidance. I also thank Mr. Yang for his guidance on this model.

References

[1] S. Beckert, D. Küster, Therapy decision support based on recommender system methods. Journal of healthcare engineering, vol. 68, pp. 1-11, 2017.

[2] M. Balvert, G. Patoulidis, A Patti, et al. A Drug Recommendation System (Dr.S) for Cancer Cell Lines. Social Science Electronic Publishing. 2019, 10(3):27-35.

[3] ZHOU Xiling.Research on medicine recommendation algorithm based on collaborative filtering.Hunan: Changsha University of Science & Technology, 2017.(in Chinese)

[4] T. Mikolov, K. Chen, G. Corrado, et al. Efficient Estimation of Word Representations in Vector Space. Computer Science, 2013.

[5] R. Johnson, T. Zhang. Deep Pyramid Convolutional Neural Networks for Text Categorization Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1, no. 2, pp. 135-142, May 2017.

[6] Y. Kim, Convolutional Neural Networks for Sentence Classification. Eprint Arxiv, 2014.

[7] R. Mutiara A. Benny, E. Wibowo. Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. Journal of Big Data,2021,8(1).

[8] C. Ronan,W. Jason,B. Léon,et al.Natural Language Processing (Almost) from Scratch.Journal of Machine Learning Research,2011,12(7).2493-2537.

[9] Lu H, Zhang M, Xu X, et al. Deep Fuzzy Hashing Network for Efficient Image Retrieval[J]. IEEE Transactions on Fuzzy Systems, 2020, PP(99):1-1.

[10] Ma C, Li X, Li Y, et al. Visual information processing for deep-sea visual monitoring system[J]. Cognitive Robotics, 2021, 1:3-11.

[11] Lu H, Zhang Y, Li Y, et al. User-Oriented Virtual Mobile Network Resource Management for Vehicle Communications[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, PP(99):1-12.

[12] Huimin LU, Yu TANG, Yi SUN. DRRS-BC: Decentralized Routing Registration System Based on Blockchain, IEEE/CAA Journal of Automatica Sinica, 2020.

[13] Sun J, Li Y. Multi-feature fusion network for road scene semantic segmentation, Computers & Electrical Engineering, vol.92, June 2021, 107155.

[14] Wang P, Wang D, X Zhang, et al. Numerical and Experimental Study on the Maneuverability of an Active Propeller Control based Wave Glider, Applied Ocean Research, 2020.

[15] H Lu, R Yang, Z Deng, et al. Chinese image captioning via fuzzy attention-based DenseNet-BiLSTM, ACM Transactions on Multimedia Computing Communications and Applications, 2021.

[16] Zheng Q , Zhu J , Tang H , et al. Generalized Label Enhancement with Sample Correlations, IEEE Transactions on Knowledge and Data Engineering, 2021.