American Journal of Biochemistry and Biotechnology

Original Research Paper

Classification Model of Traditional Chinese Medicine Prescription Decocting Duration Combining Text Convolutional Neural Network with Attention Mechanism

[1,2,4,5]Zhibiao Li, [1]Zhicheng Jiang, [3]Jianqiang Du, [6]Li Ning, [1]Huayong Zhao, [1]Yiwen Li and [4]Zhenfeng Wu

[1]*College of Computer Sciences, Jiangxi University of Chinese Medicine, Nanchang, China*
[2]*School of Pharmacy, Jiangxi University of Chinese Medicine, Nanchang, China*
[3]*School of Mathematics and Computer Sciences, Nanchang University, Nanchang, China*
[4]*Department of Technical, Ganjiang New Area Zhiyao Shanhe Technology Co., Ltd, Nanchang, Jiangxi, China*
[5]*Jiangxi Provincial Key Laboratory of Chinese Medicine Artificial Intelligence, Jiangxi University of Chinese Medicine, Nanchang, China*
[6]*Department of Nursing, Shanghai More Dental Hospital, Shanghai, China*

Article history
Received: 24-05-2024
Revised: 02-07-2024
Accepted: 24-07-2024

Corresponding Author:
Jianqiang Du
School of Mathematics and Computer Sciences, Nanchang University, Nanchang, China
Email: jianqiang_du@163.com

Abstract: The decocting time of Traditional Chinese Medicine (TCM) formulas is crucial for therapeutic effects. To better capture the important characteristics related to the decocting time of TCM formulas, a formula decocting time classification model TextCNN-attention that integrates the attention mechanism and Text Convolutional Neural Network is proposed. This model predicts the decocting time of formulas and divides them into long-term, short-term, and medium-term decocting. First, the texture information of medicinal herbs is used to expand the TCM prescriptions. The attention mechanism is then used to learn the importance of each word or subsequence in the prescription text, and the medicinal materials and medicinal texture text in the prescription are weighed. Finally, TextCNN was used to classify the prescription extension text with decocting time labels. The experimental results show that compared with the baseline model, the proposed TextCNN-attention model can better understand the expanded text information of the prescription. The experimental prediction accuracy and F1 value are improved by 1.78 and 1.87%, respectively, indicating that the TextCNN-attention model has better performance in classifying text decoction duration after the formula was extended.

Keywords: TCM Prescription, Decocting Duration, TextCNN, Attention Mechanism

Introduction

Traditional Chinese Medicine (TCM) prescriptions have made significant contributions to thousands of years of practice as a common method of treating illness and maintaining health. Decoctions, as a common form of prescriptions, have gained widespread attention because of their advantages of fast absorption and simple operation. However, the quality of decoctions is influenced by many factors, such as decocting duration, soaking time, and water addition. Particularly, the decocting time, as one of the crucial steps to ensure the efficacy of Chinese herbal medicine, will directly affect the extraction of active ingredients and the stability of medicinal effects (Zhang *et al*., 2022a). The decocting duration of TCM prescriptions is usually determined by TCM physicians based on personal clinical experiences, which may lack objective and unified standards. Different physicians may use different decocting durations for the same prescription based on their own experiences, leading to certain discrepancies in results. Only relying on personal clinical experience may result in under-decoction or over-decoction, thereby affecting the quality of decoctions (Li *et al*., 2024). Therefore, reasonable control of decocting duration reasonably is crucial for improving prescription efficacy and ensuring safety.

With the rapid development of information technology, a large amount of text data has been generated in various industries, especially common short texts. Many scholars have focused on the problem of short text classification (Xu *et al*., 2020), making it an important task in the field of natural language processing. Traditional machine learning text methods include Naive Bayes (Blanquero *et al*., 2021), K-nearest neighbors (Zhang and You, 2021), and support vector machine (Chandra and Bedi, 2021). However, these models rely heavily on manually designed features, requiring a significant amount of manpower and time. Therefore, compared with traditional machine learning models, numerous deep learning models are now being applied to text classification. Kim *et al*. (2014) proposed to apply the Text Convolutional Neural Network (TextCNN) model to text classification and use a specific Max-over-time pooling layer to select the most significant features of short texts, demonstrating the performance of TextCNN in short text classification. Lai *et al*. (2015) proposed a TextRCNN model for text classification by replacing the convolutional layers in TextCNN with RNN. Duan *et al*. (2018) proposed a hybrid deep learning structure CNN-ELM for age and gender classification, which uses CNN to extract features from input images and ELM to classify intermediate results. The effectiveness of the hybrid structure was verified on the dataset. The fusion of attention mechanisms and deep learning models for text classification problems has also been widely studied. Attention mechanisms were initially applied in the field of image recognition. When neural networks recognize an image, they dynamically score features each time they focus on the image, determining the importance of features. Since attention mechanisms can dynamically determine the importance of features, they are also been applied to text classification. Chai *et al*. (2020) used attention mechanisms to extract word information from comments to enhance CNN's focus on key information in the text. In order to improve the classification accuracy of arrhythmia, Li *et al*. (2019) used a BiLSTM model with a fused attention mechanism to learn the full sequence features of electrocardiogram signals. The classification results showed that the proposed model had good classification performance. Cheng *et al*. (2020) combined CNN and bidirectional GRU networks to extract local text features and contextual semantic information and performed sentiment classification of important words through the attention mechanism, which improved the text feature extraction ability of the model. The experiment showed that the proposed model achieved good results in sentiment classification problems. Guo *et al*. (2019a) proposed an RCNN model that integrates attention mechanisms to extract text features using recurrent neural networks and convolutional neural networks. At the same time, word-level and sentence-level attention mechanisms are introduced to enhance keywords and features for classification. Experiments on four different datasets show the good classification performance of the proposed model. In the classification of decoction time for prescriptions, the similarity between prescriptions is usually calculated for classification. Jiang *et al*. (2023) proposed a study on the decoction duration of TCM formulas based on the similarity of formulas. The Jaccard algorithm, cosine similarity algorithm, and LDA topic model algorithm were used to calculate the similarity of the composition, dosage, and function of the herbs in the formulas. The similarity of the formula in the three dimensions was fused to study the decoction duration recommendation of TCM formulas. This study adopts traditional machine learning algorithms to recommend similarity. The similarity matching based on the surface meaning of text symbols makes it easy to ignore the impact of potential dependencies between medicinal herbs on the cooking time, and the recommendation based on similarity scores cannot verify the reliability of the results. Li *et al*. (2024) developed a deep learning model that uses multi-dimensional feature weighted fusion to scientifically and effectively determine the decoction duration of TCM formulas. The model vectorizes and weights the attributes of medicinal herbs, and is then fed into the TextCNN model. Following the TextCNN pooling layer, a three-layer fully connected layer is applied. Although this experimental design captures higher-level features of medicinal herbs, it still does not pay enough attention to the attributes that influence the decoction time.

TCM prescriptions are relatively common short texts, and traditional machine learning models are insufficient in sparsity, feature representation, and inability to capture semantic representation. Attention should be paid to key TCM information and the situations where the input length of short texts for prescriptions varies and the interpretability is poor. This study has the following highlights: (1) The medicinal parts that can affect the decocting time of the prescription in TCM are used as extended text to construct a classification dataset, and the classification of decocting time is transformed into a text classification problem; (2) Attention mechanisms are used to score the importance of Chinese medicinal materials and their medicinal parts. A TextCNN prescription decocting time classification model that integrates an attention mechanism is proposed to meet the requirements of controlling prescription decocting time. The experimental results indicate that the classification of herbal decoction time makes up for the dependency relationship between medicinal herbs ignored by similar algorithms and the insufficient attention to important features that affect the decoction time of prescriptions, and achieves good experimental results.

Materials and Methods

TextCNN Model Materials

TextCNN is a text classification model based on convolutional neural networks, which can effectively learn text information by extracting local features from texts through convolutional layers pooling layers, and then classifying the texts. When CNN extracts feature information, it treats text data as one-dimensional images, and the convolutional kernels slide up and down on the

DOI: 10.3844/ajbbsp.2024.311.322

text representation matrix to extract features (Zhang, 2022). Derived from CNN, TextCNN is mainly used in text classification. It can set filter kernel ranges to combine and select local features of different sizes in the text. It includes an input layer, convolutional layer, max pooling layer, fully connected layer, and output layer. The process of the TextCNN model is shown in Fig. (1).

Input layer (Embeddings): This layer converts the text data vectors into a specified size, where each row in the matrix corresponds to a word. Assuming a sentence X has n words representing the number of rows in the matrix, the size of the sentence matrix is represented as follows:

$$X_{1:n} = X_1 \oplus X_2 \oplus X_3 \oplus \cdot \cdots \oplus X_n \tag{1}$$

where, X_n represents the i-th word in the current text, and $\oplus$ denotes the vector concatenation operation.

Convolutional layer (Conv): This layer computes the unit node matrix by convolving the input vector matrix with convolutional kernel filters. This matrix can extract features at different levels and then pass through an activation function to generate a nonlinear feature mapping matrix. The features are represented as follows:

$$C_i = f(X_{i+h-1} * W + b) \tag{2}$$

where, W is the weight matrix; f is the activation function; b is the bias parameter; X_{i+h-1} represents the matrix formed by the herbal feature vectors from the i-th row to the $i+h-1$th row, and C_i represents the resulting features obtained from convolution.

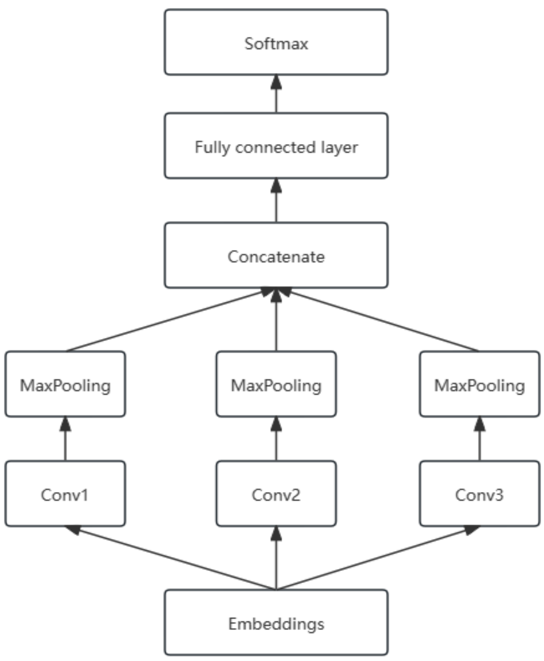

Fig. 1: Flow chart of the TextCNN model

MaxPooling layer. Assuming the dimension of the feature map after the convolution operation is L * K, where, L is the length of the convolutional sequence and K is the number of convolutional kernels. The most prominent features are extracted by taking the maximum value within the local area of the feature map. These features are preserved while discarding other features to prevent overfitting, shown as follows:

$$O_i = \max(x_{i,1}, x_{i,2}, x_{i,3}, \ldots, x_{i,j}) \tag{3}$$

where, i represents the i-th feature map of the convolutional layer; j represents the size of the pooling layer window, and O_i represents the maximum value.

Fully connected layer. To prevent overfitting, dropout is added during the forward propagation of the model to reduce the network's excessive dependence on connections, shown as follows:

$$y = r \odot x \tag{4}$$

where, x is the input vector; r is a binary random vector generated by the Bernoulli distribution with the same dimension as x, where the value of the element is 0 or 1, indicating whether each neuron is preserved $\odot$ represents multiplication operations at the element level.

Output layer (Softmax). After the fully connected layer, the label probability distribution is calculated using the softmax function, shown as follows:

$$P(y_i|x) = \frac{e^{z_i}}{\Sigma_{j=1}^{k} e^{z_j}} \tag{5}$$

where, z_i represents the i-th element of the normalized fractional vector; z_j represents the j-th element of the normalized fractional vector; K represents the number of categories, and $P(y_i|x)$ represents the probability of category y_i given the input text x.

Attention Mechanisms Materials

Attention mechanisms are a technique in neural networks that mimics cognitive attention. This mechanism can enhance the weights of certain parts of the input data in the neural network while weakening the weights of other parts, thereby focusing the network's attention on the most important part of the data (Niu *et al.*, 2021). The types of attention mechanisms can be roughly divided into soft attention mechanisms, hard attention mechanisms, and self-attention mechanisms. The core logic of attention mechanisms is to extract key information from global information and give more attention (Zhang and Liu, 2022). In addition, attention mechanisms often require the use of steps that calculate weight functions, which are selected based on the characteristics of the text being studied. Common

DOI: 10.3844/ajbbsp.2024.311.322

calculation weight functions include additive models, multiplication models, and dot product models. The calculation of the weight function determines the specific method of the model in calculating attention weights, which directly affects the model's understanding and processing ability of the input sequence. The attention mechanism consists of a query, key, and value vector, which can essentially be expressed as a query vector mapping to a series of key-value pairs (Guo *et al*., 2022; Liu and Guo, 2019). An appropriate calculation weight function should be able to dynamically allocate attention weights based on different parts of the input sequence so that the model can better process the data. The selection of weight functions for calculation is particularly important.

The additive attention model (Gao *et al*., 2021) is a common weight-scoring function in attention mechanisms. Text classification tasks can obtain the correlation between words by calculating the similarity between queries and keys, which helps to better capture semantic information in the text. The formula for the additive attention model is as follows (6):

$$Score(q,k) = w^T \tanh(W_q q + W_k k) \quad (6)$$

where, the score represents the computational function of the additive attention model. q represents the query vector; k represents the key vector; W represents a learnable weight vector; W_q and W_k represent learnable weight matrices, and Tanh represents the activation function.

The importance of the features obtained from TextCNN (Zhang *et al*., 2022b; Guo *et al*., 2019b) is evaluated. A fully connected layer is defined to map the output to the weight vector space, normalize it to obtain attention weights using the softmax function, and then apply this weight distribution to the input vector to obtain a weighted output vector, enhancing the expressive power of the model.

Research Methods

Overall Flowchart Design

This study first collected the electronic medical records of TCM from Jiangxi University of Traditional Chinese Medicine-Qihuang National Medical College from 2018-2019 after desensitization. The collected clinical TCM prescription texts issued by TCM patients are taken as the basic data. Then the composition and decocting time of TCM are extracted from the prescription. According to the records of the medicinal parts of TCM in the 2020 edition of the Chinese pharmacopoeia, including roots and rhizomes category, stems and bark category, skin category, leaf category, flower category, whole grass category, fruits and seeds category, minerals category, and other category, a dataset of medicinal parts of tcm is formed. finally, according to the dataset of medicinal parts, TCM prescriptions are expanded and vectorized. The expanded dataset is inputted into the TextCNN_attention model f to automatically extract features. The decocting time of TCM prescriptions is divided into three time periods: Label 1 (Short-term decocting time): 3-20 min; label 2 (Medium-term decocting time): 21-35 min, and label 3 (Long-term decocting time): 36-60 min. This can provide personalized decocting time plans for clinical TCM prescriptions and for assisting decision-making. The overall algorithm process of the model is shown in Fig. (2).

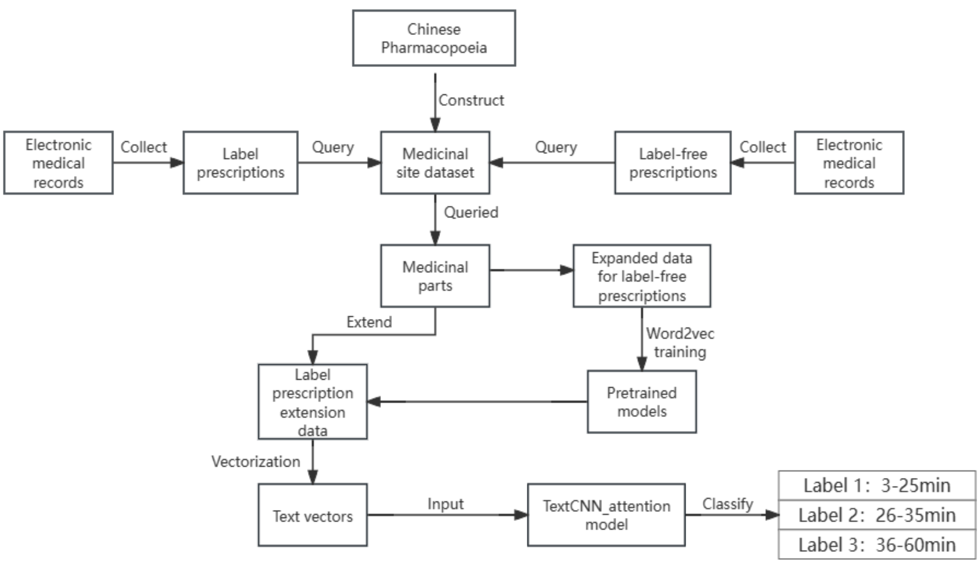

Fig. 2: Overall algorithm flowchart

Classification Standard Setting

Due to the study of the decocting time of the prescription, it was found that the growth of the prescription during the decoction process is within a certain range, and there was no accurate time point. Based on the description of the decocting time in the 2020 edition of the Chinese Pharmacopoeia, the decocting time is divided into three intervals: Prescriptions with a duration of 3-20 min, prescriptions with a duration of 21-35 min, and prescriptions with a duration of 36-60 min (Chinese Pharmacopoeia Commission, 2020).

Construction of Prescription Text and Medicinal Part Text

Electronic cases involve a wide range of data, including prescription, dosage, processing method, patient basic information, and symptoms. After data cleaning, the composition information of the prescription machine is extracted, as shown in Table (1). After data sorting and cleaning, 41550 formulas were used as an Unlabeled Prescriptions Dataset (UPD), and 14578 formulas were used as a Labeled Prescriptions Dataset (LPD). The UPD was used as the data for training the pre-trained model using word2vec in subsequent experiments, while the LPD was used as the data for training and testing model performance in subsequent experiments. The construction of prescription text is shown in Table (1). Because the prescription text belongs to the short text type and the texture of Chinese medicinal materials can directly affect the decocting time of the prescription, the attributes of the medicinal parts of Chinese medicine will be used as an extended text for subsequent prescription text decocting time classification experiments. The medicinal parts of medicinal materials can be roughly divided into 10 categories, including "Roots and Rhizomes ", and "skin". The text construction of medicinal parts is shown in Table (2).

TextCNN_Attention Model

TCM prescription texts are composed of some independent traditional Chinese medicines. From the perspective of the composition of the prescription, the text of TCM can be seen as composed of the smallest semantic unit - TCM. For example, the text of TCM "Japanese Honysuckle Flower Bud, Weeping Forsythia Fruit, Blackberrglily Rhizome, Wenchow Turmeric Root Tuber" can be divided into four types: "Japanese Honysuckle Flower Bud", "Weeping Forsythia Fruit", "Blackberrglily Rhizome", and "Wenchow Turmeric Root Tuber". Each single TCM is the smallest semantic unit. The TCM features represented by each semantic unit in the prescription have their similarities and differences, and they are combined to achieve new functional effects. It is necessary to map TCM texts to high-dimensional space vector representations so that they can understand the inherent meaning between TCM and use attention mechanisms to judge their contribution. From the perspective of text length, TCM prescription texts often consist of multiple traditional Chinese medicines with obvious short text features, making it necessary to expand their medicinal parts. This study adopts the TextCNN-attention model. It is divided into four modules: Convolution, pooling, attention, and fully connected, for the classification of the decocting time of TCM texts. The structure of the TextCNN attention model is shown in Fig. (3).

Table 1: Prescription text construction

Data set	Prescription composition	Usage
UPD	Chinese ephedrs herb, indianbueadtuckahoe, common oyster shell, liquorice root...	Pre-training
LPD	Baikal skullcap root, redpaeoniae trichocarpae, common jujube, liquorice root, ternate pinellia...	Experimental training and testing

Table 2: Text construction of medicinal parts

Medicinal parts	Medicine
Roots and rhizomes category	Ternate pinellia, pilose asiabell root, ginseng root…
Stems and bark category	Tambac, osewood heartwood, songaria *Cynomorium* herb…
Skin category	Eucommia bark, heaven ailanthus root bark, cassia bark…
Leaf category	Eucommia ulmoides oliv, hindu lotus leaf, ginkgo leaf…
Flower category	Clove, japanese honysuckle flower bud, cockscomb flower…
Whole grass category	Hairyvein agrimonia herb, capillary wormwood herb, wild mint herb…
Fruits and seeds category	Purging croton seed, ginkgo seed, Chinese arborvilae seed…
Minerals category	Alum, magnetite, mica schist, talc…
Animal category	Chinese blistering beetle, turtle carapace, dried toad venom…
Another category	Synthetic borneol, Chinese caterpillar *Fungus*, aloes…

DOI: 10.3844/ajbbsp.2024.311.322

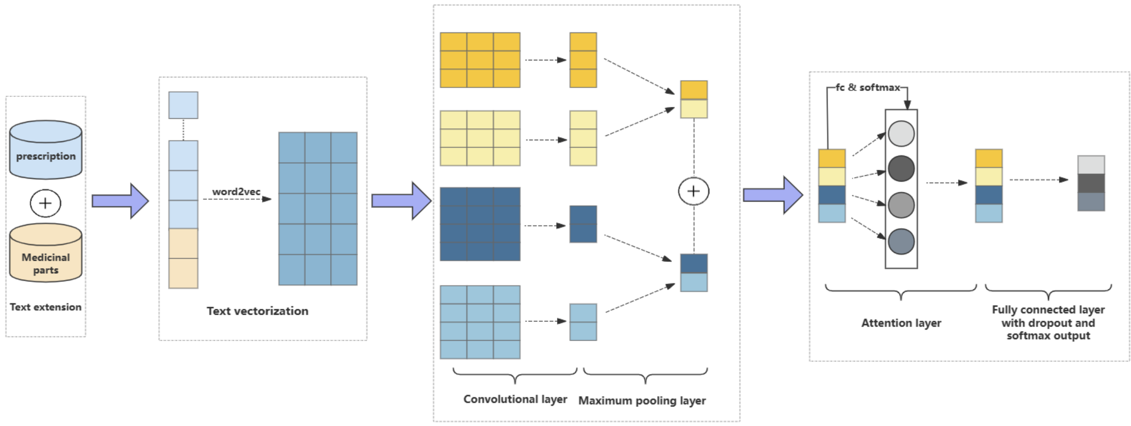

Fig. 3: TextCNN attention model structure diagram

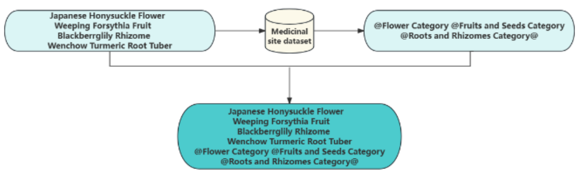

Fig. 4: Text extension example figure

Training pre-trained models: The core idea of Word2Vec (Jang *et al*., 2020; Luan *et al*., 2015) is based on the distribution assumption, which states that "words with similar context have similar semantics". There are mainly two main model structures: The Skip-gram model and the CBOW model. The Skip Gram model is given a central word and aims to predict other words that may appear in the context window of that central word. Through this approach, the model learns the distributed representation of words, making words with similar semantics closer in the vector space. The CBOW model, on the other hand, is the opposite of the Skip-gram model. The CBOW model predicts the central word given a word in a context window. Similarly, by training the CBOW model, a distributed representation of the word can be obtained. In TCM prescription texts, the occurrence rate of some traditional Chinese medicines is relatively low. Therefore, Skip Gram is used to train word vectors and map unlabeled prescription extension texts to a low dimensional dense vector space to capture semantic similarity and correlation between words, achieving effective language representation learning.

Text extension: First, based on the composition of TCM contained in the prescription text, a matching query is performed through the dataset of medicinal parts of TCM, so that the medicinal parts contained in the prescription text are concatenated to the end of the prescription text, forming an extended prescription text. For example, the original prescription text is "Japanese Honeysuckle Flower, Weeping Forsythia Fruit, blackberrglily rhizome, wenchow turmeric root tuber", and the corresponding medicinal parts "flower category, fruits and seeds category, roots and rhizomes category" are queried through the medicinal parts dataset. Finally, the two texts are concatenated together to form "Japanese honysuckle flower, weepingforsythia fruit, blackberrglily rhizome, wenchow turmeric root tuber, flower category, fruits and seeds category, roots and rhizomes category". The text expansion is shown in Fig. (4).

Word embedding layer: The input extended prescription text is transformed into a word embedding representation through a pre-trained model using word2vec. The pre-trained model plays a role in avoiding overfitting problems caused by a small dataset and has better performance and better word embedding representation to capture semantic relationships between words.

Convolutional layer: The words are embedded into the convolutional layer representing the input of the model, through which the convolutional kernel is slid on the input text to obtain a feature map that reflects the local features in the input text that match the convolutional kernel. The *tanh* activation function is used to ensure the output is between -1 and 1. Compared with the ReLU function, its output has a zero-centered characteristic, which helps to reduce the gradient vanishing problems. The *tanh* function is smooth near the origin, and the characteristic formulas are shown as follows:

$$out = tanh\big(conv(x)\big) + b \tag{7}$$

$$tanh(y) = \frac{e^{y}-e^{-y}}{e^{y}+e^{-y}} \tag{8}$$

where, x is the embedded representation of the input word; $conv$ represents the convolution operation; y represents the convolutional output result; $tanh$ represents the *S*-type hyperbolic tangent activation function, and b is the bias parameter.

Maximum pooling layer: The maximum pooling layer is then used to reduce the dimensionality of the feature map and obtain more important features. After that, the eigenvalues obtained from each convolution kernel are combined into a feature vector.

Attention mechanism layer: The main attention mechanisms are divided into soft attention mechanism, hard attention mechanism, and self-attention mechanism. However, the self-attention mechanism allows models to interact and exchange information between different positions in the same sequence. By comparing each element in the sequence with other elements, attention weights for each element are generated to process sequence data. The Chinese medicine prescription text that needs to be processed mainly consists of data from one to twenty Chinese medicinal materials and their compositions. The swapping of positions between Chinese medicinal materials does not affect the decocting time of the prescription. The hard attention mechanism does not use the Softmax function when calculating attention weights, but instead selects one or a few of the most important positions through some mechanisms (such as sampling or threshold), making it more suitable for processing long sequences or complex tasks.

The research objective text is a TCM prescription that has significant differences in the composition, taste, and quantity of TCM, as well as significant variations in length. The soft attention mechanism uses the Softmax function to normalize attention and dynamically focus on different parts of the input, so as to obtain a relatively smooth attention distribution. Therefore, the soft attention mechanism is embedded in the maximum pooling layer of the TextCNN model. The process of the soft attention mechanism is shown in Fig. (5).

The dimensions of words in a text are often low, and the interaction between words is essential for understanding the semantics and context of the text. The choice of weight function calculation will directly affect the effectiveness of the model. The dot product attention and multiplication attention are more in line with the characteristics of high-dimensional data. Compared with dot product attention and multiplication attention, additive attention can calculate the similarity between queries and keywords through a fully connected layer to better capture semantic information in text classification tasks and achieve better interaction between words.

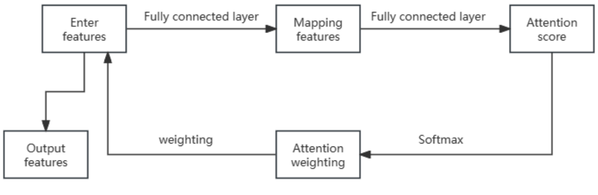

Fig. 5: Soft attention mechanism flowchart

After inputting the obtained feature vectors into the attention mechanism layer, the input features are first linearly transformed by defining a fully connected layer, and the obtained results are normalized by the softmax function to obtain attention weights. Each element is converted into a probability value that represents the attention weight of the input feature at the corresponding position. The input features are then multiplied with attention weights. The function formulas are shown as follows:

$$z = x \cdot W + b \tag{9}$$

$$a = softmax(z) \tag{10}$$

$$y = x \odot a \tag{11}$$

where, x is the output vector of the maximum pooling layer; b is the bias vector; *Softmax* is the normalization function, and $\odot$ is a symbol for the element-by-element product.

Fully connected layer: Finally, the output of the attention mechanism is input into the fully connected layer for classification.

Results and Discussion

Experimental Data Setting

The dataset extracts the required prescription texts from the desensitized TCM electronic medical record data of Qihuang Chinese medicine academy, Jiangxi University of Chinese Medicine from 2018-2019. The required prescription texts are extracted from it, and LPD is divided into training and testing sets in a ratio of nearly 10:1 for subsequent experiments. The decocting time of the data corresponds to three categories: Labels 1-3, as shown in Table (3).

Evaluating Indicator

The accuracy and F1 are used as the evaluation indicators of the model in the experiment. The definition formula is shown as follows:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{12}$$

The calculation of F1 requires precision and recall. F1 refers to the equilibrium value when the precision equals recall, shown as follows:

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

$$Recall = \frac{TP}{TP+FN} \tag{14}$$

Table 3: Experimental dataset

Decocting time min	Text category	Train set	Test set
3-20	Lable 1: Short-term decocting time	4496	411
21-35	Lable 2: Medium-term decocting time	5259	478
36-60	Lable 3: Long-term decocting time	3275	293

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} \tag{15}$$

where, *TP* represents the number of true examples with a positive prediction result; *TN* represents the true counterexample; *FP* represents the number of negative samples and positive prediction results, and *FN* is the number of positive samples and negative prediction results.

Experimental Environment Parameters and Associated Pseudocode

Experimental hyperparameter settings: The training device uses CUDA, with a random dropout probability of 0.6 and epochs set to 40. During the training process, the batch size is set to 125; the pad size for each text is set to 25; the learning rate is set to 1e-3; the convolution kernel size is (2-4), and the number of convolution kernels is 128. The main function of the code is to obtain a word.

Embedding representation of the input text, obtain text features through a convolutional layer and a max pooling layer and then create a fully connected layer through a defined attention mechanism that maps the input features to the feature space. The Softmax function is used to calculate the attention weight of the input features, and finally, the output of the attention mechanism is input into the fully connected layer for classification. The relevant pseudocodes are shown in Table (4).

Experimental Results

There are 6 types of deep learning models used for comparison:

1) BiLSTM (Liu and Guo, 2019): The advantage of a network with a unique bidirectional memory function lies in its ability to simultaneously consider the pre - and post-relationships in text sequences, capture long-term dependencies, and adapt to input texts of different lengths, which enables the model to understand text content more comprehensively
2) BiLSTM_attention (Kavianpour *et al*., 2023; Lin *et al*., 2022; Yu *et al*., 2023): A network with an attention mechanism and unique bidirectional memory function
3) DPCNN (Yu *et al*., 2021): The advantage of a deep separable convolutional neural network with a deep pyramid structure is that it effectively captures local and global information of the text through a pyramid-shaped convolutional structure while reducing the
4) number of parameters and computational complexity. The problems of gradient vanishing and information loss have been effectively solved through residual connections and pooling operations.
5) DPCNN_attention (Li and Ning, 2020; Zhang *et al*., 2023): A deep separable convolutional neural network with an attention mechanism and deep pyramid mechanism
6) TextCNN: A neural network with multiple convolutional kernels to obtain features of different sizes
7) TextCNN_attention: A neural network with an attention mechanism and multiple convolutional kernels to obtain features of different sizes

This experiment is macroscopically divided into two categories: Unexpanded Prescription Texts (UPT) classification, and Expanded Prescription Texts (EPT) classification. The experimental results of the expanded prescription text are optimal, with the highest accuracy and F1 value (in bold font), as shown in Table (5).

In the UPT classification experiment based on the baseline model, the accuracy and F1 of TextCNN can reach 76.24-76.14% respectively, which is higher than the classification of baseline models DPCNN, BiLSTM, and TextCNN.

The reason why TextCNN has a higher classification effect on decoction time than DPCNN and BiLSTM is that each medicinal herb in the Chinese medicine prescription text may represent an important local feature. The classification effect of decoction time largely depends on these features. TextCNN uses multiple convolutional kernels of different sizes, which can accurately capture local features of prescriptions through the sliding of convolutional kernel blocks, without blurring TCM information or establishing unimportant TCM relationships. DPCNN adopts residual connections and pyramid structures. Although progressive convolution and pooling operations can capture deeper features, for short texts like TCM prescriptions, the initial specific medicinal material information becomes very abstract in the final feature map, and the specific medicinal material and dosage information is blurred, resulting in poor classification performance.

Table 4: Pseudocodes related to TextCNN + attention mechanism

```
class SoftAttentionLayer(nn.Module): //Defining attention
1: out = self.embedding(x[0]) //Get word embeddings
2: out = out.unsqueeze(1) //Add dimension adaptation to convolutional layer input
3: conv_out = [F.tanh(conv(out)).squeeze(3) for conv in self.convs] //Convolutional layer output and activation
4: conv_out = [F.max_pool1d(c, c.size(2)).squeeze(2) for c in conv_out] //Maximum pooling
5: conv_out = torch.cat(conv_out, 1) //Splicing outputs of convolutional kernels of different sizes
6: conv_out = self.attention(conv_out) //Applying attention mechanisms
7: out = self.dropout(conv_out) //Dropout
8: out = self.fc(out) //Full Link Layer
```

Table 5: Experimental results

Models	Experimental results of UPT		Experimental results of EPT	
	Accuracy %	F1 %	Accuracy %	F1 %
BiLSTM	73.34	73.45	75.89	75.92
BiLSTM_ attention	75.90	75.75	77.33	77.31
DPCNN	72.65	72.82	71.07	71.59
DPCNN_ attention	73.20	73.32	73.27	73.39
TextCNN	76.24	76.14	76.73	76.65
TextCNN_	77.07	77.00	78.51	78.52

BiLSTM iteratively captures information along the entire sequence, introducing redundant information when dealing with dependency relationships in prescription texts (For example, the relationship between Japanese Honysuckle Flower and WeepingForsythia Fruitis not important, but BiLSTM will try to establish these relationships), thereby affecting classification performance. Therefore, TextCNN compensates for the limitations of DPCNN and BiLSTM. The experimental results of the baseline model on UPT data are shown in Figs. (6-7).

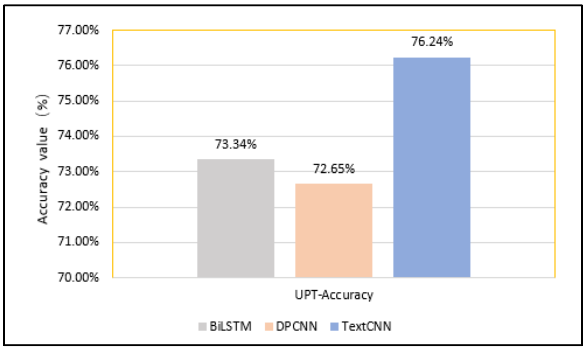

Fig. 6: Accuracy Analysis of Baseline Model Experiment Based on UPT

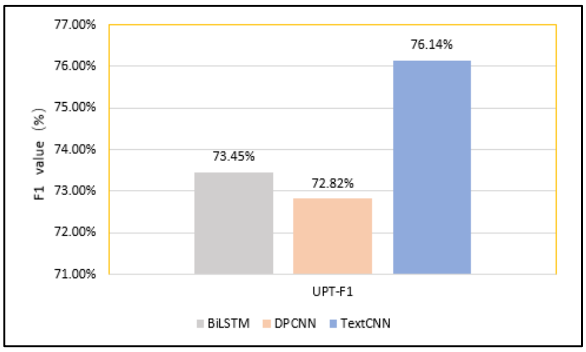

Fig. 7: Analysis of F1 values in baseline model experiments based on UPT

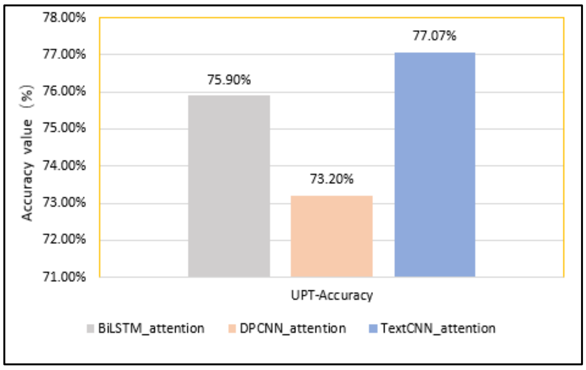

Fig. 8: Analysis of the accuracy of baseline model experiments based on EPT

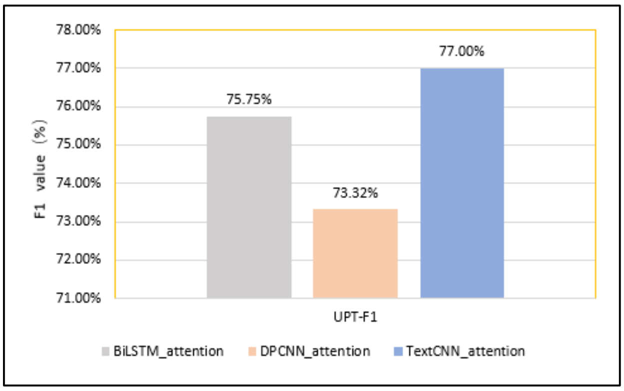

Fig. 9: Analysis of F1 values in baseline model experiments based on EPT

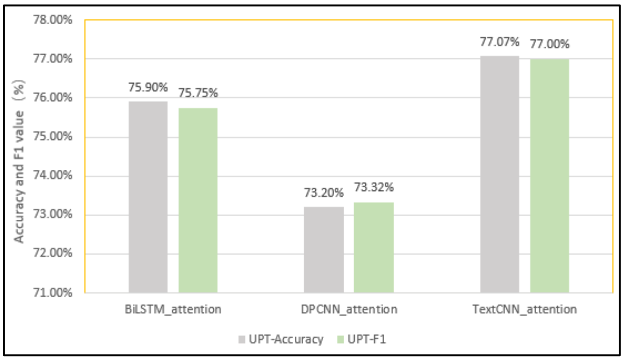

Fig. 10: Experimental analysis of fusion attention mechanism model based onUPT

The EPT classification experiment based on the baseline model shows that the accuracy and F1 value are improved compared to the UPT experiment. The EPT-based experiment aims to compensate for the short text of the prescription as much as possible in the data. The medicinal parts of the prescription are concatenated as text extensions into the unexpanded prescription text, providing more features that can affect the decocting time of the prescription. (For example, Talc and Indigowoad Leaf, the former belongs to the mineral category, and their impact on the decocting time should be greater than that of Indigowoad Leaf belonging to the leaf category). On the basis of understanding the combination and interaction between TCM prescriptions and medicinal herbs, the model can increase the semantic correlation between medicinal herbs and their compositions, helping the model achieve better classification results. The experimental results are shown in Fig. (7). The results show that after expanding the prescription text, the accuracy and F1 value of the model are improved, reaching 77.07-77.00%, respectively. The experimental results of the baseline model on EPT data are shown in Figs. (8-9).

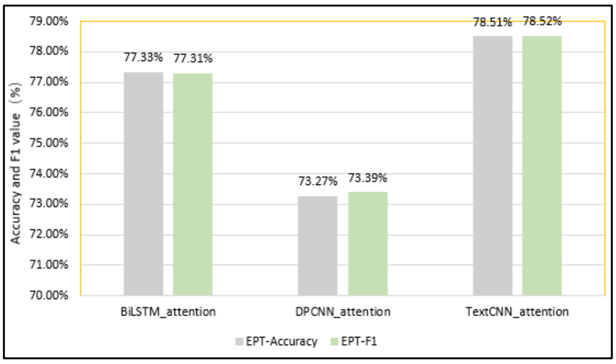

Fig. 11: Experimental analysis of fusion attention mechanism model based on EPT

This study incorporates attention mechanisms into TextCNN, DPCNN, and BiLSTM models. The experimental results show that among the three baseline models, the TextCNN_attention model cannot only achieve the best classification performance on UPT but also on EPT. The experimental results of TextCNN_attention on EPT are better than those on UPT. On the EPT data, the accuracy and F1 values are the highest, reaching 78.51-78.52%, as shown in Figs. (10-11).

Experimental analysis of UPT, a baseline model using attention mechanism, in which TextCNN_attention integrates attention mechanism into the maximum pooling layer of TextCNN. On the basis that TextCNN does not condense and fuzz TCM information and does not establish unnecessary TCM relationships, an attention mechanism is used to score the importance of the pooled TCM information features, so as to increase the interpretability of the model and improve classification performance. After incorporating the attention mechanism into the last convolutional layer of DPCNN, although the attention mechanism is added to score the importance of TCM information, it cannot handle the problem of blurred TCM information caused by progressive convolution and pooling operations. BiLSTM-attention integrates the attention mechanism into the last LSTM layer of BiLSTM. Similarly, although adding an attention mechanism can score the importance of TCM information, it cannot handle the problem of establishing unnecessary relationships between TCM materials caused by BiLSTM spreading along the entire prescription text. Therefore, in the UPT experiment, TextCNN_attention can achieve good classification performance. The experimental results of TextCNN_attention on EPT are better than those on UPT. This may be that EPT data has richer prescription text information, and can provide more important features that affect the decocting time.

Conclusion

The text classification methods in deep learning are applied to the field of TCM decoction duration and a TCM decoction duration classification model TextCNN_attention that integrates TextCNN and the attention mechanism is proposed. The text of TCM prescriptions and the medicinal parts of TCM are analyzed, and the decocting times are divided into three categories: Short-term decoction, medium-term decoction, and long-term decoction, corresponding to 3-20, 21-35 and 36-60 min. First, word2vec vectorization technology is used to obtain vectors of Chinese medicine names and medicinal parts for learning Chinese medicine features. Then, the convolutional layer and max pooling layer in TextCNN are used to extract text vector features. Finally, attention is normalized by the Softmax function of the soft attention mechanism, which can dynamically focus on different parts of the input during processing, so as to achieve a smooth attention distribution and improve the expressive power of the model. This study innovates in the field of prescription decoction duration by integrating attention mechanisms into TextCNN and expanding the prescription texts. The accuracy and F1 value of the comparative experiment are higher than common baseline models, verifying the effectiveness of the TextCNN-attention model in the text classification of TCM decoction duration. It provides doctors and patients with a reference value for prescription decoction duration and meets the needs of TCM decoction. However, this research method has two shortcomings. First, the proposed classification model overly relies on the pooled features when using attention mechanisms, and the pooling layer will result in the loss of some medicinal material-related features, making it unable to fully utilize the fine-grained features of the original prescription text. The second is to use the medicinal part attributes of medicinal herbs as an extension of the formula text, but it still cannot compensate for the short attributes of the formula text. In the future, it is necessary to address the issue of missing features in the attention mechanism of herbal text classification. In addition, research on the decoction time of traditional Chinese medicine formulas should focus on feature expansion. It can expand the TCM attributes that affect the duration of decoction, enrich the semantic representation of prescription texts (such as the efficacy and weight of TCM), and provide more features for the classification of decoction time of TCM, thereby improving the classification effect.

Acknowledgment

The author would like to express sincere gratitude to all those who contributed to this research.

Funding Information

National natural science foundation of China (82260988), science and technology plan of Jiangxi provincial department of educational (GJJ180656), the

DOI: 10.3844/ajbbsp.2024.311.322

key discipline construction of Jiangxi University of traditional Chinese medicine (2023jzzdxk025), and Research project of Ganjiang new area Zhiyao Shanhe technology Co., Ltd (ZYSH2023A002).

Author's Contributions

Zhibiao Li and Zhicheng Jiang: Designed and performed the experiments, analyzed the data, and prepared the paper.

Jianqiang Du, Li Ning and Huayong Zhao: Designed the experiments and revised the manuscript.

Yiwen Li and Zhenfeng Wu: Participated in collecting the materials related to the experiment.

Ethics

The authors declare their responsibility for any ethical issues that may arise after the publication of this manuscript.

Conflict of Interest

The authors declare that they have no competing interests. The corresponding author affirms that all of the authors have read and approved the manuscript.

References

Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable Selection for Naïve Bayes Classification. *Computers & Operations Research*, *135*, 105456. https://doi.org/10.1016/j.cor.2021.105456

Chai, Y. M., Yun, W. L., Wang, L. M., & Liu, Z. (2020). A Cross-Domain Recommendation Model Based on Dual Attention Mechanism and Transfer Learning. *Chinese Journal of Computers*, *43*(10), 1924–1942.

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their Application in Image Classification. *International Journal of Information Technology*, *13*(5), 1–11. https://doi.org/10.1007/s41870-017-0080-1

Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., & Zhong, L. (2020). Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU with Attention Mechanism. *IEEE Access*, *8*, 134964–134975. https://doi.org/10.1109/access.2020.3005823

Chinese Pharmacopoeia Commission. (2020). *Pharmacopoeia of the People's Republic of China 2020: Part 1*, China Medical Science Press, Beijing, pp: 155-168. ISBN-10: 978-7-5214-1574-2.

Duan, M., Li, K., Yang, C., & Li, K. (2018). A Hybrid Deep Learning CNN–ELM for Age and Gender Classification. *Neurocomputing*, *275*, 448–461. https://doi.org/10.1016/j.neucom.2017.08.062

Gao, C., Ye, H., Cao, F., Wen, C., Zhang, Q., & Zhang, F. (2021). Multiscale Fused Network with Additive Channel–Spatial Attention for Image Segmentation. *Knowledge-Based Systems*, *214*, 106754. https://doi.org/10.1016/j.knosys.2021.106754

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Zheng Ning, Jiang, Peng Tao, Mu, T. J., Zhang, S. H., Martin, Ralph R., Cheng, M. M., & Hu, S. M. (2022). Attention Mechanisms in Computer Vision: A survey. *Computational Visual Media*, *8*(3), 331–368. https://doi.org/10.1007/s41095-022-0271-y

Guo, B., Zhang, C., Liu, J., & Ma, X. (2019a). Improving Text Classification with Weighted Word Embeddings Via a Multi-Channel TextCNN Model. *Neurocomputing*, *363*, 366–374. https://doi.org/10.1016/j.neucom.2019.07.052

Guo, X., Zhang, H., Yang, H., Xu, L., & Ye, Z. (2019b). A Single Attention-Based Combination of CNN and RNN for Relation Classification. *IEEE Access*, *7*, 12467–12475. https://doi.org/10.1109/access.2019.2891770

Jang, B., Kim, M., Harerimana, G., Kang, S., & Kim, J. W. (2020). Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. *Applied Sciences*, *10*(17), 5841. https://doi.org/10.3390/app10175841

Jiang, M., Li, Z., Zhao, H., Shuai, D., & Liu, M. (2023). Research on a Recommendation System for Traditional Chinese Medicine Compound Decoction Plans Based on Classic Famous Decoction Methods (Original in Chinese). *Modern Information Technology*, *7*(05), 98-101,105. https://doi.org/10.19850/j.cnki.2096-4706.2023.05.05.023

Kavianpour, P., Kavianpour, M., Jahani, E., & Ramezani, A. (2023). A CNN-BiLSTM Model with Attention Mechanism for Earthquake Prediction. *The Journal of Supercomputing*, *79*(17), 19194–19226. https://doi.org/10.1007/s11227-023-05369-y

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *ArXiv*, arXiv.1408.5882. https://doi.org/10.48550/arXiv.1408.5882

Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent Convolutional Neural Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*. Proceedings of the AAAI Conference on Artificial Intelligence. https://doi.org/10.1609/aaai.v29i1.9513

Li, R., Zhang, X., Dai, H., Zhou, B., & Wang, Z. (2019). Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network. *IEEE Access*, *7*, 109870–109883. https://doi.org/10.1109/access.2019.2933473

Li, X., & Ning, H. (2020). Deep Pyramid Convolutional Neural Network Integrated with Self-attention Mechanism and Highway Network for Text Classification. *Journal of Physics: Conference Series*, *1642*(1), 012008. https://doi.org/10.1088/1742-6596/1642/1/012008

Li, Z., Zhao, H., Zhu, G., Du, J., Wu, Z., Jiang, Z., & Li, Y. (2024). Classification Method of Traditional Chinese Medicine Compound Decoction Duration Based on Multi-Dimensional Feature Weighted Fusion. *Computer Methods in Biomechanics and Biomedical Engineering*, 1–15. https://doi.org/10.1080/10255842.2024.2302225

Lin, Y., Chen, K., Zhang, X., Tan, B., & Lu, Q. (2022). Forecasting Crude Oil Futures Prices Using BiLSTM-Attention-CNN Model with Wavelet Transform. *Applied Soft Computing*, *130*, 109723. https://doi.org/10.1016/j.asoc.2022.109723

Liu, G., & Guo, J. (2019). Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. *Neurocomputing*, *337*, 325–338. https://doi.org/10.1016/j.neucom.2019.01.078

Luan, Y., Watanabe, S., & Harsham, B. (2015). Efficient Learning for Spoken Language Understanding Tasks with Word Embedding Based Pre-Training. *Interspeech 2015*, 1398–1402. https://doi.org/10.21437/interspeech.2015-56

Niu, Z., Zhong, G., & Yu, H. (2021). A Review on the Attention Mechanism of Deep Learning. *Neurocomputing*, *452*, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091

Xu, J., Cai, Y., Wu, X., Lei, X., Huang, Q., Leung, H., & Li, Q. (2020). Incorporating Context-Relevant Concepts into Convolutional Neural Networks for Short Text Classification. *Neurocomputing*, *386*, 42–53. https://doi.org/10.1016/j.neucom.2019.08.080

Yu, S., Liu, D., Zhang, Y., Zhao, S., & Wang, W. (2021). DPTCN: A Novel Deep CNN Model for Short Text Classification. *Journal of Intelligent & Fuzzy Systems*, *41*(6), 7093–7100. https://doi.org/10.3233/jifs-210970

Yu, M., Niu, D., Wang, K., Du, R., Yu, X., Sun, L., & Wang, F. (2023). Short-Term Photovoltaic Power Point-Interval Forecasting Based on Double-Layer Decomposition and WOA-BiLSTM-Attention and Considering Weather Classification. *Energy*, *275*, 127348. https://doi.org/10.1016/j.energy.2023.127348

Zhang, C., Guo, R., Ma, X., Kuai, X., & He, B. (2022a). W-TextCNN: A TextCNN Model with Weighted Word Embeddings for Chinese Address Pattern Classification. *Computers, Environment and Urban Systems*, *95*, 101819. https://doi.org/10.1016/j.compenvurbsys.2022.101819

Zhang, J., Jiang, Y., Wu, S., Li, X., Luo, H., & Yin, S. (2022b). Prediction of Remaining Useful Life Based on Bidirectional Gated Recurrent Unit with Temporal Self-Attention Mechanism. *Reliability Engineering & System Safety*, *221*, 108297. https://doi.org/10.1016/j.ress.2021.108297

Zhang, M. J., Pang, J., Cai, J., Huo, Y., Yang, C., & Xiong, H. (2023). DPCNN-Based Models for Text Classification. *2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, 363–368. https://doi.org/10.1109/cscloud-edgecom58631.2023.00068

Zhang, S. (2022). Challenges in KNN Classification. *IEEE Transactions on Knowledge and Data Engineering*, *34*(10), 4663–4675. https://doi.org/10.1109/tkde.2021.3049250

Zhang, T., & You, F. (2021). Research on Short Text Classification Based on TextCNN. *Journal of Physics: Conference Series*, *1757*(1), 012092. https://doi.org/10.1088/1742-6596/1757/1/012092

Zhang, X., & Liu, H. (2022). Effects of Different Decocting Methods and Time on the Treatment of Constipation Caused by Yang Deficiency with Compatibility of Processed Fuzi with Rhei Radix et Rhizoma and Its Influence on the Heart (In Chinese). *China Rational Drug Use*, *19*(07), 87–93. https://doi.org/10.3969/j.issn.2096-3327.2022.07.015