

## Research Article

# Text Feature Extraction and Representation of Chinese Mental Verbs Based on Deep Learning

Yongxue Wang<sup>1</sup> and Tee Boon Chuan <sup>2</sup>

<sup>1</sup>Chinese Studies, Institute of Studies, Universiti Tunku Abdul Rahman, Kajang 4300, Malaysia

<sup>2</sup>Institute of Chinese Studies, Universiti Tunku Abdul Rahman, Kajang 4300, Malaysia

Correspondence should be addressed to Tee Boon Chuan; b20160803111@stu.ccsu.edu.cn

Received 28 July 2022; Revised 19 August 2022; Accepted 25 August 2022; Published 8 September 2022

Academic Editor: Lianhui Li

Copyright © 2022 Yongxue Wang and Tee Boon Chuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are some problems in feature extraction and representation of Chinese mental verbs, such as low accuracy and low efficiency. In order to further improve the computational efficiency and accuracy of Chinese mental verb text, based on deep learning theory, activation function and damage function were used to optimize the original model. Considering the calculation method of model gradient, the optimization model describing the characteristics of Chinese mental verbs is finally obtained. The model can be used to analyze the variation of the characteristic parameters of Chinese verbs and the method of representation. Finally, the model error is analyzed by the method of comparative verification. Relevant studies show that the number of outputs and output results corresponding to softmax function will influence the test results of the model. By comparing the curves, it can be seen that the curve corresponding to the output number has an obvious increasing trend, while the corresponding output result curve has an opposite changing trend. The linear and nonlinear characteristics of the two curves are obvious. The real value of the mean square error function shows a change of linear increase, while the corresponding output value shows a change trend of gradual decline, which indicates that the two kinds of data have different influences on the model under related algorithms. It can be seen from the error data that the gradual increase of independent variables will improve the accuracy of the test results. Five different Chinese mental verb parameters have different manifestations in the deep learning model: among them, declarative verbs fluctuate in a small range and have little corresponding influence. However, the fluctuation of nondeclarative verbs and positive and negative declarative verbs is relatively small, and the curve is relatively stable. Negative verbs have a positive influence on the test output. Double negative verbs have negative effects. Finally, the accuracy of the model is verified by calculating the difference between experimental data and model data. This research can provide theoretical support and model verification method for the application of deep learning model in other fields of Chinese language.

## 1. Introduction

Deep learning models have been widely applied in artificial intelligence and other fields and also have obvious application prospects in face recognition [1], information security [2], molecular microscopy [3], retinopathy monitoring [4], and other fields. In view of the low calculation accuracy and slow calculation efficiency of partial differential equation in the calculation process, based on deep learning theory, fuzzy analysis method was adopted to extract the features of calculus calculation process, so as to obtain the periodic characteristic parameters. Through damage estimation of

characteristic parameters, an optimized deep learning model was obtained [5]. This model can provide new research thinking for the solution of calculus equation and can further improve the solution accuracy of the model and further compress the solution time of the model. Finally, the accuracy of the model was verified by comparing experimental data with model data. Through verification and analysis, it can be seen that relevant theories based on deep learning model can provide reference for solving calculus equation. The analysis and prediction of cancer have always been a difficult medical problem. There were some problems in the process of cancer analysis, such as backward analysis

methods and complicated cancer lesions. In order to further improve and optimize the tissue framework in the process of cancer treatment, artificial intelligence technology was used to analyze the original model based on deep learning model, so as to obtain a new optimization model [6]. This model mainly uses different types of activation functions to analyze the framework of treatment, so as to extract the required data and characteristic indicators. The model algorithm was used to analyze these indexes and find out the changing rules of different indexes. Finally, the deficiencies of existing data can be further optimized according to these calculation and research rules, so as to obtain the final optimization model. Through experimental calculation, different frames and types of the model can be analyzed, and finally, the accuracy of the model can be verified by experimental data.

The above studies mainly started from the fields of industry and medical treatment and failed to provide better solutions to the problems existing in the analysis of characteristic parameters of Chinese mental verbs. Based on deep learning theory, this paper uses activation function and damage equation to analyze and study the text features of Chinese mental verbs. Model gradient analysis was used to further characterize the text features of Chinese verbs so as to find the corresponding representation methods. Finally, the analysis model of Chinese mental verbs based on deep learning theory was obtained. This model can provide research ideas for the analysis of the text characteristics of Chinese mental verbs and use the method of index calculation to analyze the different indicators of Chinese. Finally, the error analysis method was used to verify the model, and the results show that the model has good advantages and application prospects. This study can provide research ideas for the textual analysis of Chinese mental verbs.

## 2. Theoretical Basis of Deep Learning

Based on the relevant theories of deep learning model, this paper mainly uses activation function and damage function to extract the text features of Chinese mental verbs, so as to get the change rules of Chinese mental verbs under different indicators [7, 8]. In order to introduce the change rules of different indicators into the original deep learning model, model gradient algorithm is adopted to classify and analyze the change rules of different indicators [9, 10]. The common points of indicators were found and introduced into the original model combined with algorithm rules. The optimized deep learning model was obtained by revising the original model. The calculation of model can calculate the text features of Chinese mental verbs, and the calculation results can well reflect the actual change rules of verbs. The ultimate goal is for machines to be able to learn analytically, like humans, and to recognize data such as text, images, and sound. Deep learning is a complex machine learning algorithm that has achieved far more results in speech and image recognition than previous related technologies.

Deep learning model has been widely used in feature extraction of Chinese mental verbs [11, 12]. In order to further illustrate the application process of this model in the field of Chinese mental verbs, the training process of the

corresponding deep learning model is obtained through summary analysis, as shown in Figure 1. The coordinates of the corresponding text features of Chinese mental verbs should be extracted first, and then the multipath errors are analyzed as training samples. By importing the model samples into the sample database, and then setting the relevant parameters, the corresponding analysis model of Chinese mental verb text features based on deep learning theory is obtained, and the model is further built. In order to further illustrate the accuracy of the model, the training accuracy of the model should be identified first. If the model does not meet the requirements, the corresponding sample should be imported into the sample input for circulation. If it meets the requirements, the grid of the model should be saved and then analyzed in the output grid under the action of coordinate sequence, so as to finally get the corresponding error mining.

*2.1. Activation Function.* Each layer of neural network contains several neurons. Through repeated training and learning, the weight value of neuron connection is adjusted, so as to obtain data processing ability [13, 14]. Neuron is the basic element of neural network. Neuron structure consists of input signal, sum, threshold judgment, activation function, and output. Input values of neurons are transmitted by input signals of other neurons through their respective connection weights, and finally, output is generated through activation function processing.  $x_i$  is the input signal transmitted by other neurons,  $w_i$  is the corresponding connection weight of other neurons,  $\theta$  is the threshold, and  $y$  is the output, which is output by the following formula:

$$y = f\left(\sum_{i=1}^m w_i x_i - \theta\right), \quad (1)$$

where  $f(x)$  is the activation function of the neuron.

In the activation function, different parameters have different effects on the output value [15, 16]. The variation rules of output value and activation function under the action of different neurons were further explained, and the calculation indexes of different neurons were obtained as shown in Figure 2. When the number of neurons exceeds 17, the curve increases rapidly, showing the characteristics of nonlinear change, and the increase range of the curve is relatively large. It can be seen from the corresponding activation function that it shows a downward change with the increase of neurons, and then shows a change trend of approximately linear increase with the increase of neurons. The curve then rapidly increased to the maximum value and remained stable, and then decreased rapidly under the action of higher neurons, with a relatively large fluctuation range. Through the curve changes of the two, it can be seen that the overall change is relatively obvious, with opposite change characteristics, which can reflect the actual rule of parameter change.

The activation function in neurons is usually nonlinear. Since the input and output of neurons are linearly correlated, the neuron has more powerful performance and can process

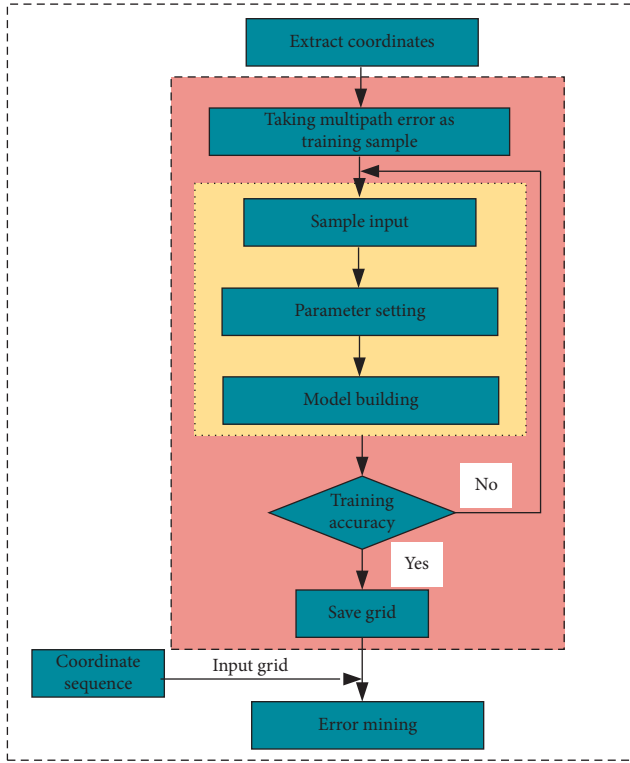


FIGURE 1: Flow chart of deep learning training.

complex feature information through the nonlinear operation of activation function. The choice of activation function changes with the further research. Common activation functions include the following:

- (1) Sigmoid function, its formula is as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \quad (2)$$

The characteristic of this function is that its output value varies between (0, 1) and can be used for dichotomies. In information science, Sigmoid function is often used as threshold function of neural network due to its properties of singleton and inverse singleton. The main disadvantages of the Sigmoid function include the following: when the input value is large, the function is not sensitive to its change perception, which easily leads to the situation that the gradient is 0 in the back propagation, resulting in slow update of weight in network training.

- (2) Tan *h* function, its formula is as follows:

$$\tan hx = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \quad (3)$$

This function can be regarded as magnifying and translating the Sigmoid function, and its output value varies between [-1, 1], and the output mean value is around 0. Compared with Sigmoid function, Tan *h* function performs better, has the effect of data center, and can accelerate network training, but it

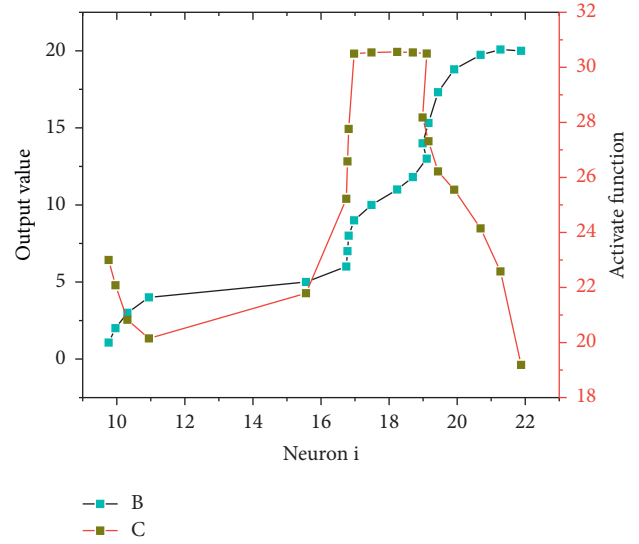


FIGURE 2: Different neuron computation index.

also has the same gradient disappearance. Moreover, the calculated data corresponding to Tan *h* function have more extensive output results, which makes the research scope of Tan *h* function more extensive.

- (3) ReLU function, whose formula is expressed as follows:

$$f(x) = \max(0, x). \quad (4)$$

This function is the most commonly used activation function in neural networks and can accelerate the convergence of network models. Compared with Sigmoid function and Tan *h* function, ReLU function has the advantage that when the input value is greater than 0, it will not tend to saturation with its change. However, there are also obvious disadvantages: the constant gradient of some neurons is 0, so that the corresponding parameters can never be adjusted, namely, the phenomenon of neuron necrosis.

- (4) Leaky ReLU function, whose formula can be expressed as follows:

$$f(x) = \max(ax, x). \quad (5)$$

By introducing constant *a*, this function solves the neuron failure phenomenon of ReLU function in the region whose input value is less than 0 and effectively reduces the disappearance of gradient.

Leaky ReLU function is an optimization of ReLU function in a linear interval. To further illustrate the accuracy of the optimization, the Leaky ReLU function and ReLU function as well as the calculation results between the parameters have been obtained through calculation and analysis, as shown in Figure 3. The test data show a two-stage variation trend with the increase of independent variables: in the first stage, the curve gradually drops to the lowest

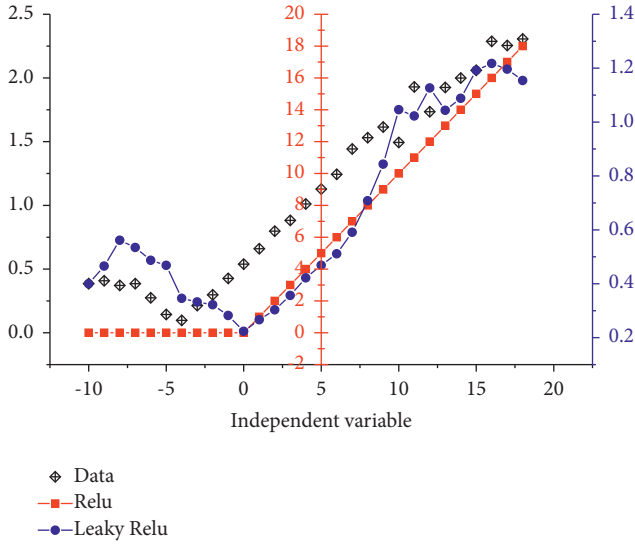


FIGURE 3: ReLU function optimization results.

point. Then in the second stage, the curve gradually increases to the maximum value, and the increment is basically the same, indicating that the linear characteristics are relatively obvious. The corresponding traditional ReLU function tends to zero in the first stage, which cannot better describe the decline and change of the test data in the first stage. In the second stage, ReLU function showed a trend of linear increase, but it could not reflect the test data well in terms of specific values and could only be consistent with the test data in the trend. Leaky ReLU function can not only explain the change of its linear decline in the first stage but also describe the change trend and specific data in the second stage relatively well. Therefore, the optimized Leaky ReLU function can better reflect the change rule of experimental data.

(5) Softmax function, its formula is as follows:

$$\text{Softmax}(x) = \frac{\exp(x)}{\sum_{i=1}^I \exp(x)}, \quad (6)$$

where  $x$  is the output of the  $i$ th neuron and  $I$  is the total number of output neurons of this layer. This function can realize feature mapping and is often used in multiclassification problems.

In softmax function, the number of neuron input and output result will affect the output value. In order to further explain the variation rule of output number and output result under different output time, the time variation curve of softmax function under different output time was calculated as shown in Figure 4. It can be seen from the changes in the figure that the number of output and the change of output result have an opposite trend. Firstly, it can be seen from the number of output that it shows a trend of slow increase with the increase of output time, and then in a higher output time, the curve shows a gradual increase.

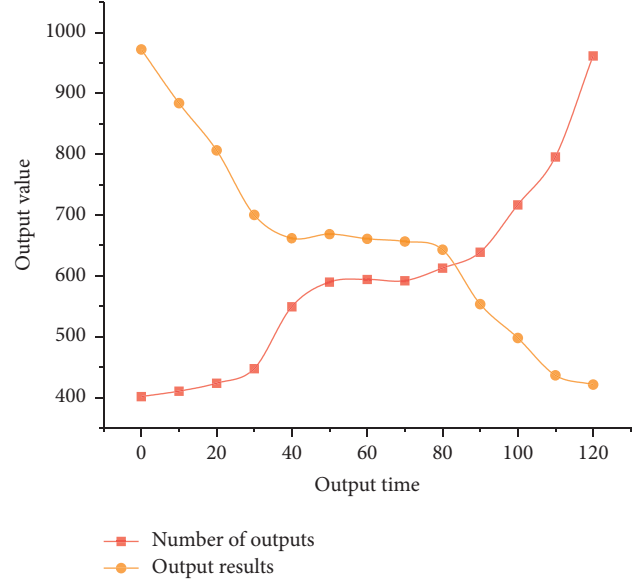


FIGURE 4: Change value of softmax function at different output time.

When the curve increases to the local maximum value, the curve gradually tends to be stable. At a higher output time, the corresponding output value of the output number curve increases rapidly to the maximum value, and the range of the curve is relatively large. The output results show a linear decreasing trend under the action of lower output time, and then the curve gradually tends to be stable. Under the action of higher output time, the corresponding output results tend to be stable. This shows that the number of outputs increases linearly while the output results decrease linearly.

2.2. *Loss Function.* Loss function is used to calculate the error between the output value and the real value during model training, which can reflect the training effect of the model in real time and provide direction for model training optimization [17, 18]. The smaller the loss function, the better the robustness of the model. Common regression loss functions include mean square error loss and mean absolute error loss. Common classification loss functions include piecewise loss, binary loss, and exponential loss [19, 20].

(1) Mean square error loss function: this function is the most commonly used loss function in regression tasks, and its formula is expressed as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2, \quad (7)$$

where  $N$  is the number of true values,  $y_i$  is the output value of the model, and  $\bar{y}_i$  is the true value. The loss function does not consider the direction of the error, but only the average value of the error, which can better reflect the error degree between the output value and the real value. Because the variation relationship between the real value and the output value is considered, the loss function is easy to

calculate the gradient. However, since the loss function assumes that the error between the output value and the real value of the model meets the Gaussian distribution, the error of outliers will be too large, resulting in poor robustness.

It can be seen that the real value and the output value have different influences on the specific data of the model by calculating the mean-square error loss function in the loss function. In order to further explain the change rule between the real value and the output value and the corresponding change of the error, the curve of the error value of the mean-square error loss function was obtained through calculation as shown in Figure 5. The true value shows a gradually increasing trend with the increase of the independent variable, and the change amount of the true value decreases is relatively consistent, indicating that its linear characteristics are obvious. With the increase of independent variables, the corresponding output results show a trend of gradual decline, and the slope of decline is constant, which indicates that the real value and the corresponding output value have obvious linear characteristics. The corresponding error is obtained through calculation, and the corresponding error curve is basically maintained at a higher level under a lower independent variable. When the independent variable increases gradually, the curve drops rapidly to the lowest point and tends to be stable gradually under the action of higher independent variable. This indicates that the independent variable will have a certain influence on the test results and also indicates that there is a certain error between the real value and the output value.

- (2) Average absolute error loss function, which is given by the absolute error value between the output value and the true value, and its formula is expressed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}_i|, \quad (8)$$

where  $(y_j - \bar{y}_j)$  is the learning error. In the process of gradient descent, the loss value of the loss function will be positive or negative 1, which is not conducive to the training of the model. Although the loss function assumes that the error between the output value of the model and the real value satisfies the Laplace distribution, the Laplace distribution is more robust to the processing of outliers.

- (3) The piecewise loss function means that the loss is 0 when the output value is equal to the true value and 1 when the output value is unequal. The formula is expressed as follows:

$$F(y, f(x)) = \begin{cases} 0, & y = f(x), \\ 1, & y \neq f(x). \end{cases} \quad (9)$$

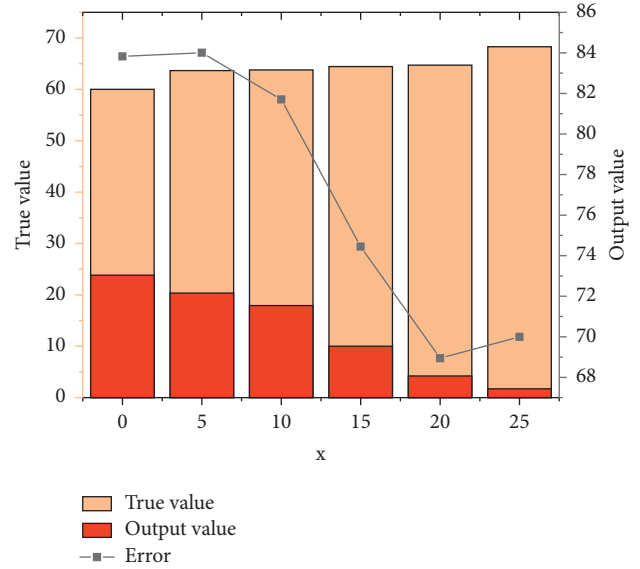


FIGURE 5: Curve of error of mean square error loss function.

The loss function directly reflects the number of classification errors.

- (4) Binary loss function is the most commonly used loss function in model classification loss calculation. It is often used in binary classification problems and can effectively reflect the degree of deviation from the true value of the output value. Its formula is expressed as follows:

$$E(y_i, \bar{y}_i) = - \sum_{i=1}^N p(y_i) \log p(y_i) - (1 - p(\bar{y}_i)) \log (1 - p(y_i)), \quad (10)$$

where  $N$  is the number of true values,  $P(y_i)$  is the output probability distribution value of the model, and  $P(\bar{y}_i)$  is the true probability distribution value. In the process of gradient descent, the binary loss function has a good learning characteristic of fast update with large error weight and slow update with small error weight. Considering the probability distribution of model parameters, the calculation results of binary loss function are more consistent with the characteristics of data distribution.

- (5) Exponential loss function is often used in model algorithms and is very sensitive to noise and outliers. Its formula is expressed as follows:

$$Z = \exp[-yf(x)]. \quad (11)$$

It can be seen from the above analysis that the classified loss function includes three different functions: piecewise function, binary function, and exponential function. There are differences in the calculation formulas of the three different functions, indicating that there are some differences in the corresponding calculation results. In order to further

explain the change rules of different specific functions in the classification loss function, the change curves of the classification loss function are obtained through calculation, as shown in Figure 6. First of all, it can be seen from the test data that the curve first decreases to the lowest point and then gradually tends to be stable. However, when the independent variable gradually increases, the test data show a linear increase in the change law, and its change range is relatively large. As can be seen from the piecewise loss function, when the independent variable is less than 20, the corresponding output value is zero, while when the independent variable is more than 20, the corresponding curve increases. Then, the corresponding output value shows a fluctuation trend and finally gradually tends to decline. The overall variation range of the curve is relatively small. Through the binary loss function, it can be seen that the curve drops first, then fluctuates, and finally gradually rises in three stages. The corresponding range of variation is relatively small, overall between 0.2 and 0.5. It can be seen from the exponential damage function that it has typical exponential function change characteristics. Under the action of small independent variables, the curve shows a rapid downward trend, and then the curve gradually tends to be stable. Under the action of higher independent variable, the curve decreases gradually. This shows that three different loss functions have different typical characteristics, and it is necessary to comprehensively consider two or three functions to accurately describe and analyze the test data.

**2.3. Model Gradient Descent.** The training process of neural network includes forward propagation and back propagation [21, 22]. Backpropagation is the core part of network training to realize the dynamic adjustment of network training parameters. Forward propagation propagates data from the shallow layer to the deep layer of the neural network until the output layer is cut off [23, 24]. During the calculation of gradient descent, the neural network can obtain the error between the output value and the real value through the loss function. Back propagation is to obtain the updated weight value of neurons calculated by gradient descent method through the errors generated by forward propagation, and then use the chain derivative formula to carry out the back propagation calculation. In gradient descent, the method of partial derivative is usually used to calculate the gradient value of the current parameter. Then a learning rate  $\eta$  is introduced to control the change speed of updating weight value, and the weight updating formula is as follows:

$$W = w - \eta \frac{\partial J(w)}{\partial w}. \quad (12)$$

However, in order to update the parameter values of other layers in the neural network, it is necessary to make further use of back propagation. In this process, the back propagation uses the chain rule to calculate the gradient:

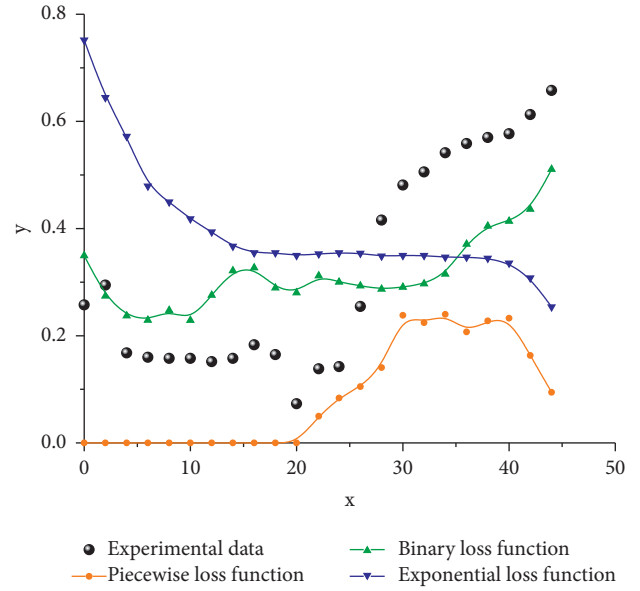


FIGURE 6: Classification loss function curve.

$$\frac{\partial J(w)}{\partial w} = \frac{\partial J^*(x)}{\partial y} \times \frac{\partial y}{\partial w}, \quad (13)$$

where  $w$  is the parameter value of the network and  $y$  is the output value of the network activation function.

Therefore, back propagation can pass the updated parameter value to the corresponding position layer by layer through the chain rule using the error function. Thus, the parameter value update formula of layer  $L$  network is as follows:

$$W = w - \eta \frac{\partial J(w)}{\partial w} = w - \eta \left[ \delta \times (x^{L-1})^T \right], \quad (14)$$

where  $x$  is the input value.

The accuracy of the model can be further illustrated by analyzing the formula of the model gradient descent and the corresponding calculation results. It can be seen that different learning rates and corresponding parameter values will have different influences on the output results of the model. In order to further illustrate the influence of learning efficiency and parameter values on test data, the model gradient curve was obtained through calculation, as shown in Figure 7. It can be seen from the curve that the learning rate curve and the parameter value curve have different change rules. Firstly, it can be seen from the test data that, with the increase of input value, the corresponding test data first increases and then decreases, and the overall fluctuation range is relatively small. When the input value of the curve gradually increases, the corresponding output results show a trend of gradual increase; when it reaches the maximum value, the corresponding test data gradually decreases with great volatility. This shows that the nonlinear characteristics of the test data are obvious. As can be seen from the learning rate curve, with the increase of input value, the corresponding curve fluctuates in a larger range first, and then gradually increases to the maximum value. As the value

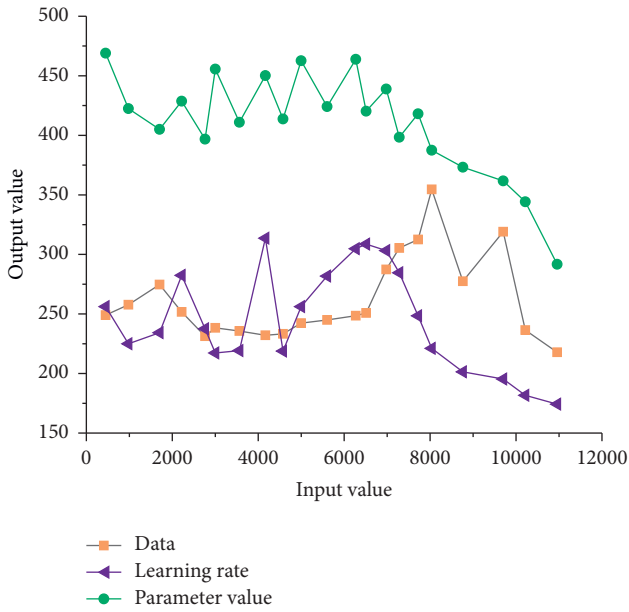


FIGURE 7: Model gradient curve.

increases, the curve drops rapidly and then flattens out. It can be seen from the variation curve of parameter values that the parameter curve fluctuates rapidly under the action of small input values, and the fluctuation range is relatively small. Then, as the input value gradually increases, the corresponding output result gradually decreases to the lowest value.

### 3. Analysis of the Characteristics of Chinese Mental Verbs Based on Deep Learning

**3.1. Text Feature Extraction of Chinese Mental Verbs.** There are many problems in the analysis of the textual features of Chinese mental verbs, mainly including the unclear description and analysis of Chinese features. In order to further analyze the characteristics of Chinese mental verb text, five different feature parameters are obtained through analysis: declarative verbs, nondeclarative verbs, positive and negative verbs, negative verbs, and double negative verbs. In order to further explain the proportion of these five verbs in actual Chinese mental verbs, the pie chart of Chinese mental verbs is obtained through statistics as shown in Figure 8. The proportion of declarative verbs is about 10%, while that of nondeclarative verbs is about 15%. The proportion of positive and negative declarative verbs is the smallest, only 7%. Negative verbs make up the most, about 43 percent, while double negatives make up about 25 percent. It can be seen from the above analysis that the proportion of negative verbs is the highest, while the proportion of positive and negative verbs is the lowest.

**3.2. Analysis of Text Features of Chinese Mental Verbs Based on Deep Learning.** Based on the deep learning model, the activation function and loss function are used to analyze the original model, so as to obtain the optimization value of the

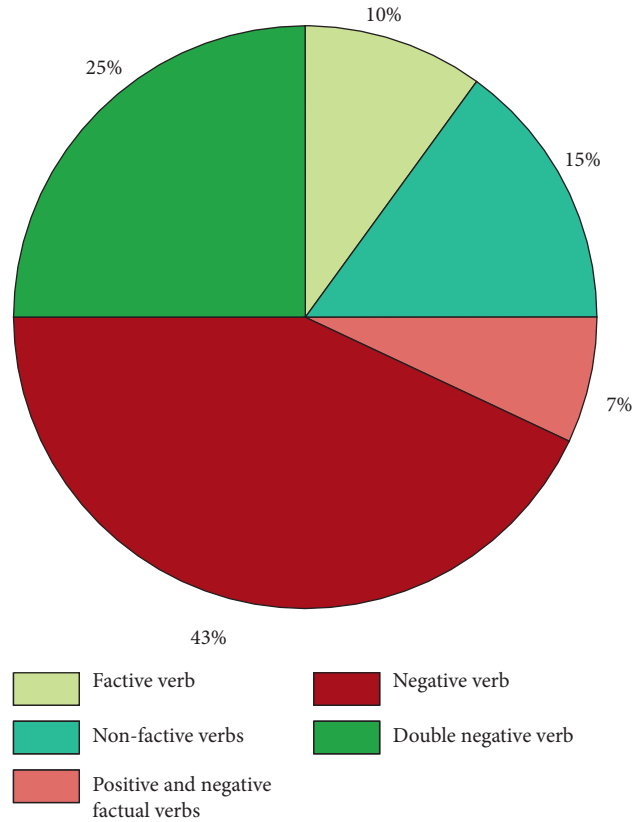


FIGURE 8: Pie chart of Chinese mental verbs.

deep learning model. The model gradient descent is introduced into the model to obtain the optimized deep learning model, which can be used for the analysis of Chinese mental verbs [25]. In order to further illustrate the application of the deep learning model in Chinese mental verbs, the flow chart of the application of the deep learning model in Chinese mental verbs is obtained through the above analysis, as shown in Figure 9. As can be seen from the computational flow chart of text feature extraction of Chinese mental verbs, the data of Chinese mental verbs are first imported into the model and then divided into different sections according to different feature parameters. Then, the critical value of the model can be obtained according to the verb feature of the calculated value. Based on the critical value, the verb feature can be further clipped and analyzed. The results of clipping and analysis are introduced into the activation function to analyze the change between the activation function and the damage function. Furthermore, the linear difference and gradient descent results are obtained to obtain the reconstruction scheme and model of Chinese mental verbs. Finally, the model calculation results and optimization results are output.

Based on the deep learning model, the activation function and damage function are used to optimize the original model, so as to obtain the optimization model for describing Chinese mental verbs. This optimization model can calculate the characteristic parameters of Chinese mental verbs, and thus obtain the histogram of the calculation results of characteristic parameters of Chinese mental

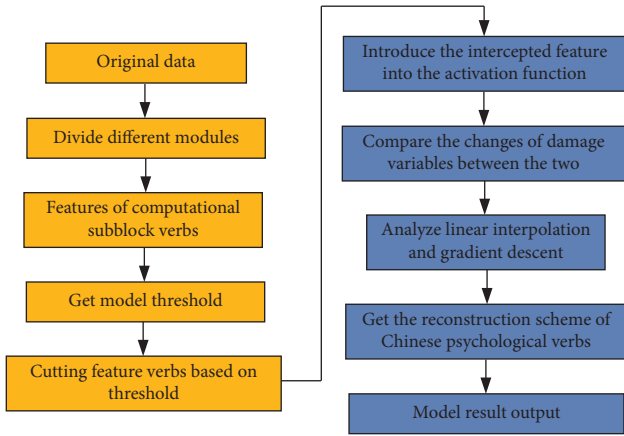


FIGURE 9: Flow chart of deep learning algorithm.

verbs, as shown in Figure 10. It can be seen from the figure that different characteristic parameters show different variation trends with different sample parameters. First of all, it can be seen from the calculation results of declarative verbs that the calculation results are relatively small and the overall range of change is relatively small. A trend in volatility in which a curve increases gradually, then drops to a low point, and then increases gradually. While the corresponding nonpredicate verbs show a linear increase trend with the gradual increase of samples. When the curve reaches the local maximum value, the curve gradually decreases to the local minimum value with the increase of samples. As can be seen from the positive and negative narrative verbs, the change of positive and negative narrative verbs gradually decreases with the increase of samples, and the change of positive and negative narrative verbs gradually tends to be stable with the further increase of samples. It shows that with a higher sample size, positive and negative verbs tend to be constant. However, it can be seen from the negative verb that its change increases linearly with the increase of samples, and the overall change range of the curve is relatively large. As can be seen from the change of double negative verbs, the curve has a high output result under the action of small specimens, while the corresponding output result drops to the lowest value rapidly with the increase of samples, and the overall change range is relatively large.

#### 4. Discussion

The optimization model based on deep learning theory can carry out targeted data analysis by considering activation function and loss function. In order to further illustrate the superiority of the calculated results, the test curve, model curve, and corresponding error curve were obtained by summarizing, as shown in Figure 11. As can be seen from the model verification curve, the test curve first presents a V-shaped change with the increase of iteration steps, and then rapidly drops to the lowest point and then carries on a small fluctuation. With the further improvement of iteration steps, the test data gradually increased to the maximum value, and the overall change range was approximately U-

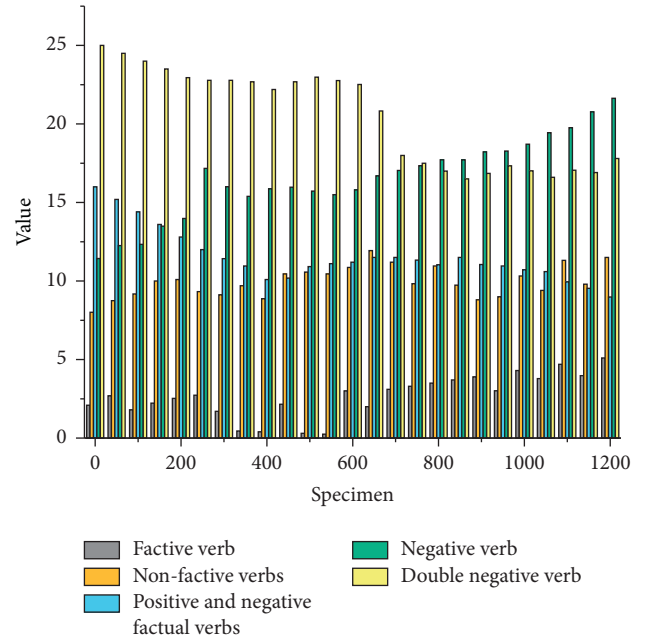


FIGURE 10: Histogram of calculation of characteristic parameters of Chinese mental verbs.

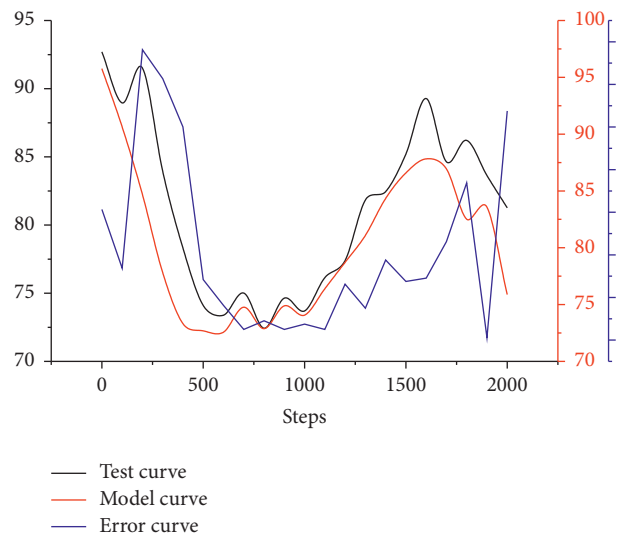


FIGURE 11: Model validation curve.

shaped. The corresponding model curve can better reflect the overall change trend of the test data, and then get the error curve through the comparison between the two. It can be seen from the error curve that the overall error range is within 7%, while with the increase of the number of test steps, the corresponding error range gradually decreases. As the sample increases, the curve is driven by increasing variation. This shows that the error is also U-shaped in a certain range. The curve has different error ranges under different iteration steps. Therefore, in practical application, it is necessary to determine the number of iterations according to the specific test data, so as to obtain the optimal calculation results.



## 5. Conclusion

- (1) It can be seen that the output value and the activation function have the same variation range through the change curves of the calculated indexes of different neurons. When the neuron is low, the output value and the corresponding activation function have the same change trend. When the neuron is high, the output value shows a rapid decline, while the corresponding activation function shows a rapid increase.
- (2) The comparison curve between Leaky ReLU function and ReLU function shows that the original model can only provide a good description of the second stage of the test data. The optimized activation function can not only explain the change rule of the first stage but also better analyze the experimental data in the key points of the second stage.
- (3) Model gradient changes mainly include learning rate and parameter values of the model. By comparing the experimental data, it can be seen that the curve shows obvious fluctuation in the first stage, among which the fluctuation of learning rate is relatively large. The corresponding parameter values change in a small range, and the curves all decline significantly in the second stage.

## Data Availability

The dataset can be obtained from the corresponding upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] X. Sun, P. Wu, and S. Hoi, "Face detection using deep learning: an improved faster RCNN approach," *Neuro-computing*, vol. 299, no. 19, pp. 42–50, 2018.
- [2] A. Rajkomar, E. Oren, K. Chen et al., "Scalable and accurate deep learning with electronic health records," *Npj Digital Medicine*, vol. 1, no. 1, pp. 18–36, 2018.
- [3] E. Nehme, L. E. Weiss, and T. Michaeli, "Deep-STORM: super-resolution single-molecule microscopy by deep learning[J]," *Optica*, vol. 5, no. 4, pp. 436–452, 2018.
- [4] M. Voets, M. Kajsa, and L. A. Bongo, "Replication study: development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]," *PLoS One*, vol. 10, no. 12, pp. 77–96, 2018.
- [5] J. Han, J. Arnulf, and E. Weinan, "Solving high-dimensional partial differential equations using deep learning[J]," *Proceedings of the National Academy of Sciences*, vol. 63, no. 12, pp. 406–423, 2018.
- [6] D. Bychkov, N. Linder, R. Turkki et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Scientific Reports*, vol. 8, no. 1, pp. 3395–3412, 2018.
- [7] T. Ben-Nun and T. Hoefler, "Demystifying parallel and distributed deep learning: an in-depth concurrency analysis," *ACM Computing Surveys*, vol. 52, no. 4, pp. 4023–4045, 2018.
- [8] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials[J]," *ACS Nano*, vol. 78, no. 42, pp. 4203–4215, 2018.
- [9] Y. Li, W. Shi, and W. W. Wasserman, "Genome-wide prediction of cis-regulatory regions using supervised deep learning methods," *BMC Bioinformatics*, vol. 19, no. 1, pp. 202–225, 2018.
- [10] D. S. Kermany, M. Goldbaum, W. Cai et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.
- [11] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: a brief review," *Computational Intelligence and Neuroscience*, vol. 2018, no. 12, pp. 1–13, 2018.
- [12] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: a survey," *Computers and Electronics in Agriculture*, vol. 147, no. 12, pp. 70–90, 2018.
- [13] H. Chen, O. Engkvist, and Y. Wang, "The rise of deep learning in drug discovery," *Drug Discovery Today*, vol. 23, no. 6, pp. 63–86, 2018.
- [14] Q. S. Zhang, S. C. Zhu, and U. O. California, "Visual interpretability for deep learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 415, no. 42, pp. 639–657, 2018.
- [15] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: a survey," *IEEE Access*, vol. 6, no. 5, pp. 14410–14430, 2018.
- [16] Y. Tom, H. Devamanyu, and P. Soujanya, "Recent trends in deep learning based natural language processing [review article]," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.
- [17] X. X. Zhu, D. Tuia, and L. Mou, "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience & Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2018.
- [18] E. Nachmani, E. Marciano, and L. Lugosch, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 85, no. 40, pp. 1–10, 2018.
- [19] L. He, K. Ota, and M. Dong, "Learning IoT in edge: deep learning for the internet of things with edge computing," *IEEE Network*, vol. 32, no. 1, pp. 96–101, 2018.
- [20] H. He, C. K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.
- [21] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, no. 26, pp. 146–157, 2018.
- [22] S. Macaluso and D. Shih, "Pulling out all the tops with computer vision and deep learning," *Journal of High Energy Physics*, vol. 18, no. 10, pp. 736–752, 2018.
- [23] A. Cocos, F. Alexander, and A. Masino, "Reply to comment on: 'deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts'," *Journal of the American Medical Informatics Association Jamia*, vol. 780, no. 120, pp. 726–758, 2019.
- [24] Z. Rui, R. Yan, and Z. Chen, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, no. 15, pp. 213–237, 2019.
- [25] R. Brent and L. Boucheron, "Deep learning to predict microscope images," *Nature Methods*, vol. 16, no. 5, pp. 882–896, 2018.