

Research Article

Deep Learning for Chinese Language Sentiment Extraction and Analysis

Zhu Zhu 

School of Education, Yunnan Open University, Kunming 650500, China

Correspondence should be addressed to Zhu Zhu; zhuzhu@ynou.edu.cn

Received 30 March 2022; Revised 20 April 2022; Accepted 26 April 2022; Published 2 June 2022

Academic Editor: Man Fai Leung

Copyright © 2022 Zhu Zhu. 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.

In recent years, vocabulary emotion processing has become immensely popular and the requirements for language emotion analysis mining and processing have become significantly abundant. The sentiment extraction and analysis work has always been very challenging; especially, the Chinese word segmentation operation is difficult to deal with effectively, the multiple combinations of implicit and explicit words make the task of sentiment analysis mining more difficult, and, in particular, the efficiency of machine analysis of language sentiment is feeble. We use some expressions and sentiment vocabulary dictionaries combined with hybrid structures and use information synergy methods to get in touch with sentiment analysis methods. We use the relevant sentiment to evaluate the explicit or implicit emotional association of the emotional connection of the vocabulary and add the unique emotional word matrix to analyze the related clustering results of the emotional words to continuously optimize and upgrade the performance, so that our sentiment analysis results are systematic in terms of efficiency and significantly improved.

1. Introduction

In recent years, the processing of language vocabulary sentiment has gradually increased and sentiment analysis mining processing is an important source of reviewing the author's opinion and attitude. We focus on the study of the time span and causal connection of emotion. Combining the dual connection of time and causality, a corresponding data model can be established and emotional data analysis can be performed on it to predict related events. Using deep learning series algorithms to analyze data units such as image networks, statistical results are obtained. The algorithm operation of the convolutional network has become the preferred method in the application of analytical images. Metrics tuning of deep learning performance has huge benefits, enabling the perfect combination of flexibility and uniformity. We use POS (part of speech) tagging and maximum entropy (ME) modeling to develop text-based sentiment detection models to explore mutual sentimental connections between words and objective transactions and perform multiple analyses of word sentiment data.

2. Introduction to the CLSTM Model

The application of the LSTM model has the following advantages: (1) It is convenient for sequential modeling work. (2) The memory function of the model is relatively powerful, and it can carry out long-term memory. (3) The model implementation scheme is relatively simple. (4) It solves some problems of long sequences, such as gradient disappearing and exploding. The CLSTM model is a neural network prediction model based on deep learning. On the model's attention mechanism, it can quickly distinguish between regular words and analysis words with emotional tendencies and compare with the traditional LSTM model. Its advantages are used in time. It is more obvious, and the classification and analysis efficiency of vocabulary is greatly improved. The CLSTM model is constructed to realize the language emotion analysis of deep learning, improve the accuracy of emotion extraction and analysis, optimize the processing and analysis results, and then realize the function of language emotion analysis [1]. The CLSTM model is shown in Figure 1.

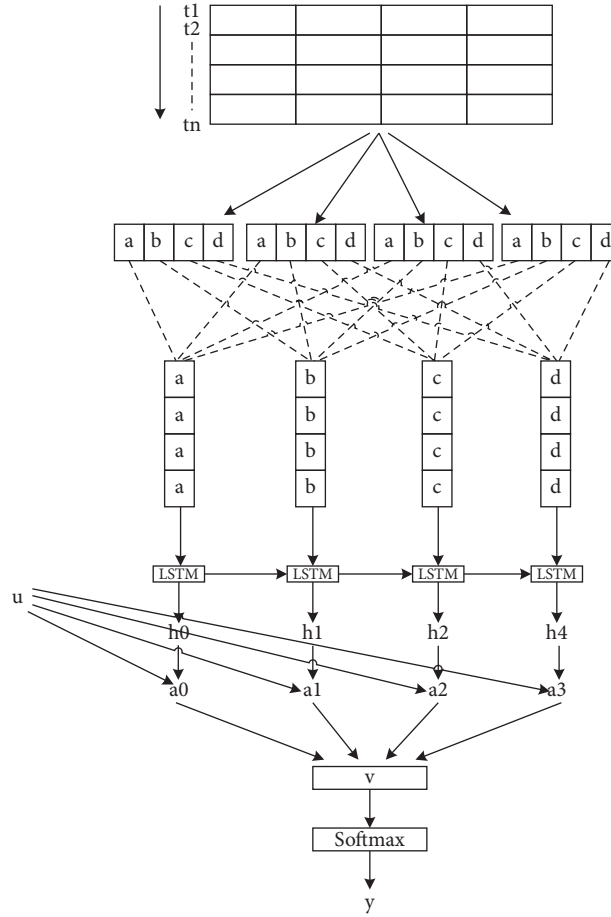


FIGURE 1: CLSTM model with attention mechanism.

2.1. Calculation of the Model. To realize the specific classification mechanism of the CLSTM algorithm, the specific implementation formula of the related algorithm is as follows:

$$T = (t_1, t_2, \dots, t_n), \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f), \quad (3)$$

$$O_t = \sigma(W_0 \cdot [h_{t-1}, X_t] + b_0), \quad (4)$$

$$\tilde{c}_t = \tan h(w_c \cdot [h_{t-1}, x_t] + b_c), \quad (5)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (6)$$

$$h_t = o_t \cdot \tan h(c_t), \quad (7)$$

$$u_t = \tan h(w_s h_t + b_s), \quad (8)$$

$$a_t = \text{soft max}(u_t^T, u_s), \quad (9)$$

$$v = \sum_t a_t h_t, \quad (10)$$

$$y = \text{soft max}(v). \quad (11)$$

2.2. Final Output of the Model. The CLSTM model combines the advantages of CNN and LSTM. CNN effectively extracts the n-gram features of the text, and LSTM has the ability to effectively capture contextual information. The CLSTM model combines these two advantages and derives its own unique advantages and uses the unique neural network data elements to effectively distinguish text data [2].

Among them, in the convolution layer, feature maps have different color information; if feature maps have the same color, it means that the convolution kernel has selected a unified operation value, and the same color will be automatically spliced to form a new vector. At this time, we can get the matrix formula as

$$w = [C_1; C_2; \dots; C_n]. \quad (12)$$

“Semicolon” represents vector splicing, that is, a value obtained by the convolution behavior of different convolution kernels in the feature map, and the dotted line part represents the generation of a new vector. A filter $f \in R^{m \times k}$ with a step size of 1 can completely get $w \in R^{(n-m+1) \times n}$; then, we should apply the new vector and gradually input it into

the LSTM, and finally, the softmax layer can output the classification probability.

In the training set, the preprocessing operation is performed, the text is vectorized, and feature extraction is performed. Finally, through the system display of the model training, a suitable model is obtained for sentiment analysis of related language texts [3]. The left-path process is the preprocessing training work. We conduct research on the training model just because we need to test the optimal results of the algorithm, so we use the optimal algorithm model to output the formula. After the left-path training model is gradually formed, we can get the right-hand training set of the standard model. Therefore, the left-hand training model must be used to determine the best model, and then the right-hand process must be carried out. The operation process is shown in Figure 2.

As mentioned earlier, CNN can be combined with LSTM to form a new CLSTM model. Here, we also call it a CLSTM model that introduces residual connections. In the past, we may have encountered a series of problems, such as loss of network parameter update data or disappearance of vocabulary gradients, not to mention that better analytical work can be achieved after concatenating the residuals. The residual connection of the CLSTM model is shown in Figure 3.

W_1 is the text representation word vector, and the output-related features of the language can be expressed as $C^i = \{C_1^i, C_2^i, \dots, C_n^i\}$; after the residual connection makes $-w_{i-1}$ and c_i as the module input, the output is $h^i = \{h_1^i, h_2^i, \dots, h_n^i\}$, and the calculation formula is as follows:

$$h_j^i = C_j^i + w_j^{i-1}, \quad (j = 1, 2, 3, \dots, n). \quad (13)$$

3. Improvement of the CLSTM Model

The specific implementation of the superimposed application of this model is shown in Figure 4.

The superposition of the algorithm is used not to randomly perform mixed operations, but to find the comprehensive performance advantages of the superposition algorithm through specific applications. In Figure 4, what we need to add is the specific operation of the hybrid operation of the CLSTM algorithm and PML and experimental steps need to be added to reflect the effectiveness and superiority of its superposition algorithm.

3.1. The PMI Model Algorithm Design. Here, we add a reference to the degree of language lexical association and use the SO-PMI algorithm to calculate the emotional trend weighted value of language vocabularies, record the language emotion comparison differences of related vocabularies, and calculate whether the emotional correlation and the degree of connection are close [4]. The formula for calculating the similarity of language vocabulary is as follows:

$$\text{PMI}(w_1, w_2) = \log_2 \left(\frac{P(w_1 \& w_2)}{P(w_1)P(w_2)} \right). \quad (14)$$

Here, $P(w_1 \& w_2)$ represents the probability of w_1 and w_2 appearing at the same time and $P(w_1)P(w_2)$ represents the probability of w_1 and w_2 appearing alone. In fact, $\text{PMI}(w_1, w_2) = \log_2(M \cdot df(w_1 \& w_2) / df(w_1)df(w_2))$ in the article is composed of the basic formula $\text{PMI}(w_1, w_2) = \log_2(p(w_1 \& w_2) / p(w_1)p(w_2))$, combined with $p(\text{word}) = df(\text{word}) / M$. In this formula, $p(w_1 \& w_2)$ represents the probability of word 1 and word 2 appearing together and $p(w_1)$ and $p(w_2)$ represent the probability of each word appearing alone. The basic idea is to use statistical methods to calculate the common probability of two words appearing in a text, and the size of the probability reflects the degree of association between the two words.

After performing the SO-PMI algorithm, we can get the emotional tendency of the language emotion of words, which can be roughly divided into neutral words, positive words, and negative words. We then add and subtract weights to determine neutral words as 0, positive words as 1, and negative words as -1 [5]. In general, if the final calculation result of the weight value is positive, the sentence can be judged to be a positive sentence; if the final calculation result of the weight value is negative, it can be judged that the sentence is a negative sentence. The algorithm can make sentiment analysis results more accurate than before in the application of word linking [6]. The implementation process is roughly as shown in Figure 5.

3.2. Introduction of the Auxiliary Algorithm Formula

3.2.1. Gradient Descent Algorithm. In this paper, the gradient descent algorithm is used simply to solve the problem of the optimal solution of the vocabulary in the current gradient, and the algorithm will explore the optimal solution of the vocabulary along the negative direction of this gradient. With the deepening of the training of the deep learning model, the difficulty level and complexity of the vocabulary will gradually deepen. Therefore, we use the negative direction to explore, so that the loss function decreases the fastest. The Taylor calculation formula of the relevant $f(x)$ is as follows:

$$f(x + a \vec{d}) = f(x_k) + ag_k^T \vec{d} + o(a), a > 0. \quad (15)$$

Here, the unit vector is \vec{d} and θ is the angle between \vec{d} and $-\vec{g}_k$; then,

$$\vec{g}_k^T \vec{d} = -\vec{g}_k \cos \theta. \quad (16)$$

When $\theta = 0$, $\vec{g}_k^T \vec{d}$ will reach the minimum value and the speed of $f(x)$ will reach the maximum value.

The descent algorithm is shown in Figure 6.

3.2.2. Sentiment Word Feature Formula Application. A piece of text is composed of many words, so we say that words are the basis of text composition [7]. In particular, the emotional vocabulary that appears in the text will greatly affect the language emotional tendency of the entire sentence. The amount of emotional vocabulary has a great effect on the appeal of the language. We supplement the emotional

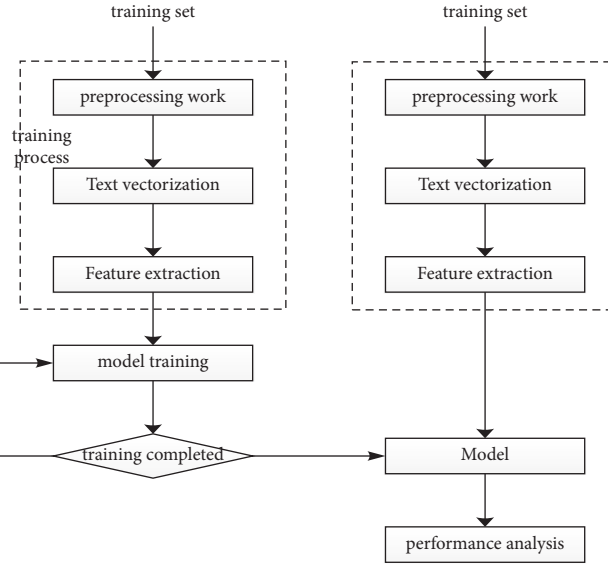


FIGURE 2: Text recognition sentiment analysis model process.

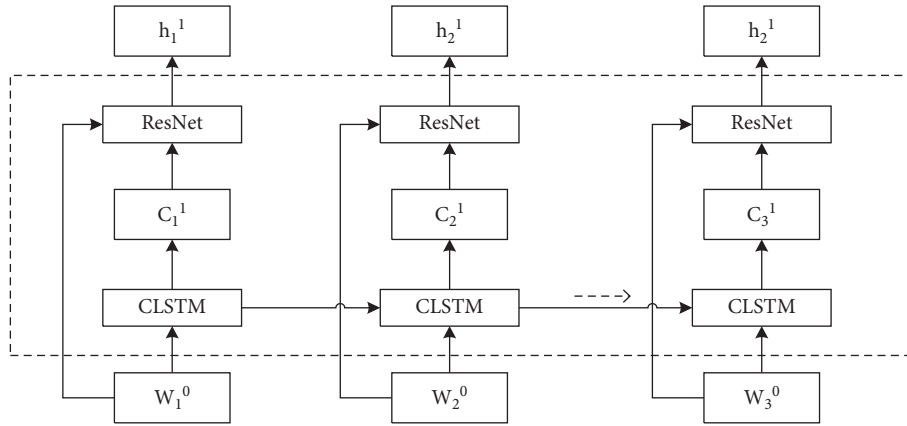


FIGURE 3: CLSTM model with residual connections.

dictionary here, which can make the emotional attribute $f_{\text{sentiment}}(s_i)$ look like the following formula:

$$f_{\text{sentiment}}(s_i) = \frac{\sum_{k=1}^m \text{estimate}(\text{word}_{i,k})}{m}. \quad (17)$$

Here, word represents vocabulary, $\text{word}_{i,k}$ is the k th word in the s_i sentence, and m represents the number of words. When $\text{estimate}(\text{word}_{i,k}) = 0$, the word is not an emotional word, and when $\text{estimate}(\text{word}_{i,k}) = 1$, the word is an emotional word. When using the relevant sentiment calculation formula, we can still add some auxiliary formulas, so on the basis of it, we have added the scoring formula calculation associated with it.

$$f_{\text{keyword}}(s_i) = \sum_{j=1}^n \text{keyword}(w_{ij}), \quad (18)$$

$$\text{keyword}(w_{ij}) = \begin{cases} 1; & w_{ij} \text{ a keyword,} \\ 0; & w_{ij} \text{ not a keyword.} \end{cases}$$

Here, w_{ij} represents the j th word in the sentence s_i ; we traverse it, and if it is a keyword, the feature score can be +1.

Of course, the analysis of all languages is inseparable from the huge language storage database. Under the relevant operations of deep learning, we use relevant equipment to remove language redundancy and extract relevant information [8]. After a series of word segmentation operations, the language is carried out. Sentiment classification, combined with the weighted value operation of the previous technology, can make a better judgment, so as to realize the extraction of the text language [9].

w_c is the central reference vocabulary of the model application, which can effectively discriminate the context content w_o and make probabilistic analysis. The relevant formulas applied by the model are as follows:

$$P(w_o|w_c) = \frac{\exp(u_o^T v_c)}{\sum_{i \in V} \exp(u_i^T v_c)}. \quad (19)$$

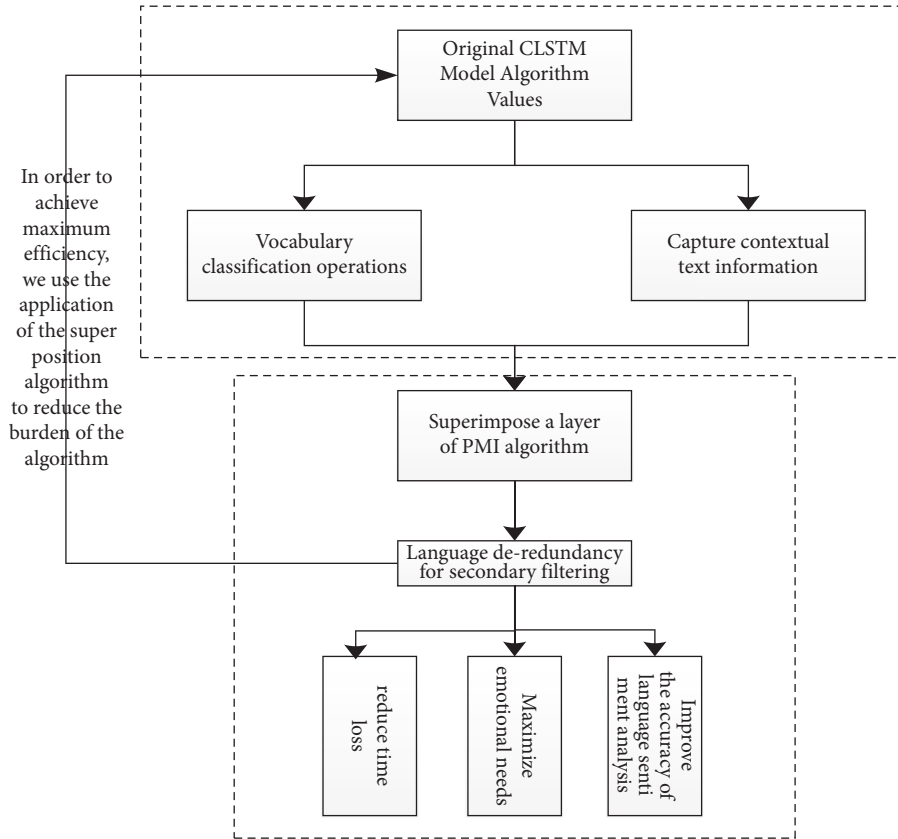


FIGURE 4: Superposition application of the CLSTM model algorithm.

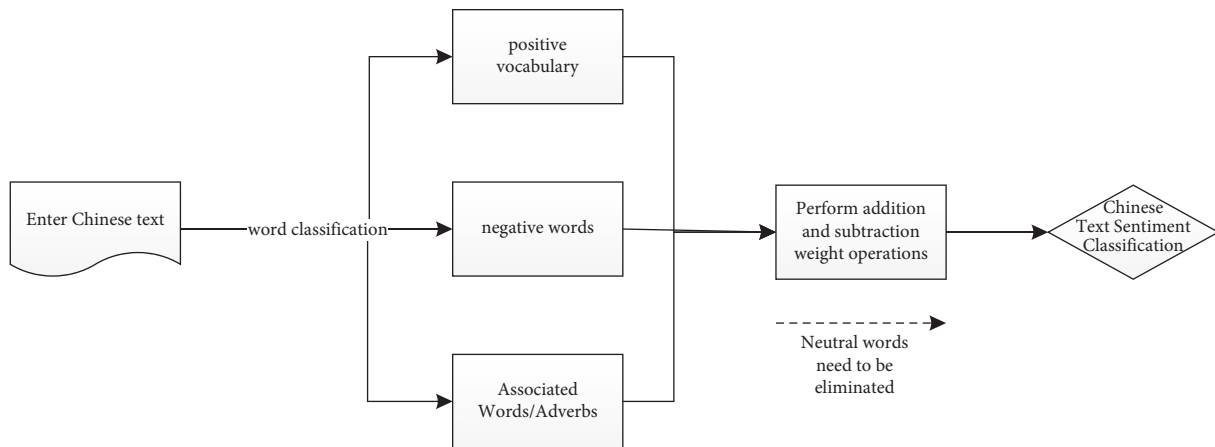


FIGURE 5: Flowchart of Chinese language emotion classification.

v_c is the center word vector, and u_i is the background word vector. We can make the corresponding flow chart, as shown in Figure 7.

Extracting and analyzing common vocabulary and selecting this formula can restore the emotional data of the original text to the greatest extent.

$$p(x_i) = \left(\sqrt{\frac{Z(x_i)}{t}} + 1 \right) \times \frac{t}{Z(x_i)} = \sqrt{\frac{t}{Z(x_i)}} + \frac{t}{Z(x_i)}. \quad (20)$$

x_i represents the original vocabulary, $Z(x_i)$ represents the probability that the sampled vocabulary is extracted

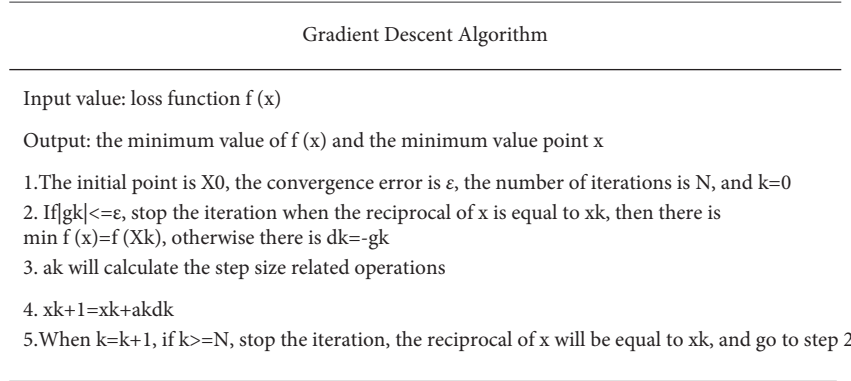


FIGURE 6: The detailed process of the gradient descent algorithm.

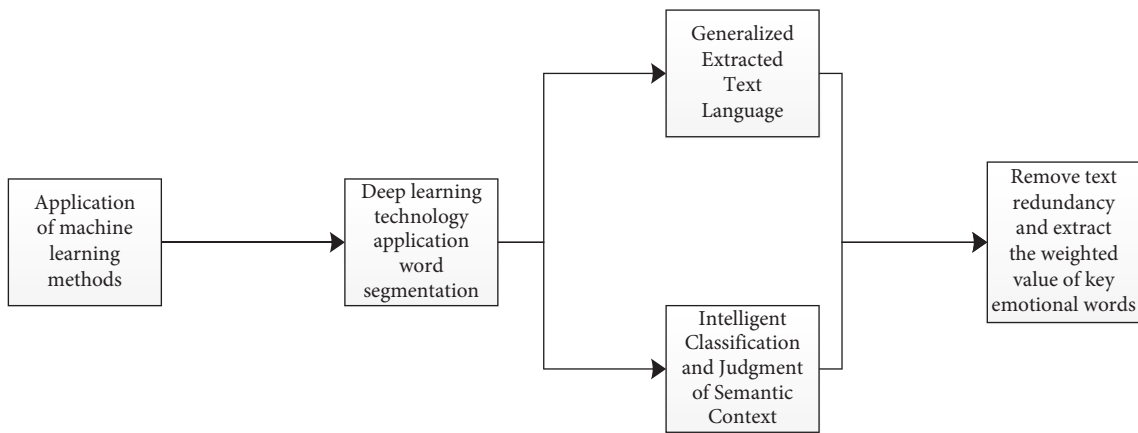


FIGURE 7: De-redundancy exclusion operation of machine learning.

from the entire data set, and t is the threshold of the sampled vocabulary, which is 0.001 by default. The larger the value unit, the greater the difference in the sampling probability. For the same vocabulary, the larger the t value, the greater the probability of the sampled value being deleted [10].

3.3. Actual SO-PMI Algorithm Operation. We can introduce relevant calculation algorithms of the probability theory. Assuming that the total number of text data is M , what we need to count is the specific number of text words, that is to say, the frequency problem in the text data set [11]. We can set it as $df(\text{word})$, which has the corresponding probability calculation formula of

$$p(\text{word}) = \frac{df(\text{word})}{M}. \quad (21)$$

Then, we can convert the SO-PMI algorithm to the formula, and the new word similarity calculation formula can be obtained as

$$\text{PMI}(w_1, w_2) = \log_2 \left(\frac{M \cdot df(w_1 \& w_2)}{df(w_1)df(w_2)} \right). \quad (22)$$

In this calculation, we use the base 2 logarithmic formula, so that the statistical result is always between 0 and 1,

so the similarity of words can be easily expressed. If the PMI value is greater than 0, it can indicate that the words have a correlation; if the PMI value is equal to 0, the words are independent of each other; if the PMI value is less than 0, the words have no correlation [12].

We now continue the introduction of the formula, and we get the calculation formula of emotional tendency. At the beginning, we need to select words with clear emotions as reference objects and calculate the PMI difference, so as to judge the difference of language emotional tendency [13]. The relevant calculation formula is as follows:

$$\text{SO-PMI}(w) = \sum_{i=1}^k \text{PMI}(\text{key} - p_i, w) - \sum_{i=1}^k \text{PMI}(\text{key} - n_i, w). \quad (23)$$

Here, $\text{key} - p_i$ represents the i th positive word in the vocabulary statistics and $\text{key} - n_i$ represents the i th derogatory word in the vocabulary statistics.

The following three judgment results for judging the Chinese language vocabulary are given as follows:

If the value obtained by SO-PMI is greater than 0, the sentiment tendency of the word w is positive.

If the value obtained by SO-PMI is equal to 0, the emotional tendency of the word w is neutral.

If the value obtained by SO-PMI is less than 0, the sentiment tendency of the word w is negative.

At the beginning of this algorithm, the first thing to do is the preprocessing of the text.

At this time, we need to extract candidate words according to the emotional language vocabulary database and judge whether the words in the text are emotional words [14]. If it has become an emotional vocabulary, the operation process can be directly skipped. If the text contains words that are not emotional words, the formula is used for calculation and evaluation.

The SO-PMI value between the word and the benchmark word is calculated. If the result is greater than zero, it can be judged as a positive value and then added to the analysis operation; if the result is less than zero, it is judged as a derogatory word and added to the analysis. After all the judgment operation steps are finished, the algorithm can be judged to be finished. Due to the diversity and complexity of the Chinese language vocabulary, we can synonymously divide the selected positive and derogatory words in consideration of the benchmark value.

Please note the following: when the target algorithm starts to execute, the corpus must be preprocessed and analyzed, then the part-of-speech judgment criteria are added after the word division and classification work, and finally, the regions are allocated [15]. This is based on the difficulty of the Chinese language. The Chinese phrase system is too large, and it is inevitable that there will be some omissions in the emotional thesaurus. Moreover, in today's society in the Internet age, the number of emotional vocabulary updates is far greater than the number of thesauri stored. Therefore, we need to design an algorithm to judge the emotion of words. The SO-PMI algorithm in the figure is one of them. The specific implementation of the SO-PMI algorithm is shown in Figure 8.

4. Simulation Experiments

4.1. Data Budget Processing. In order to realize the superiority of related operation analysis of superimposed applications, we perform superimposition processing on the SO-PML algorithm, taking 5000+ Chinese language emotion data as an example, using the following methods to deeply analyze language emotion [16]. The specific experimental methods are as follows:

Option 1: we model and cite the CLSTM model separately to analyze Chinese lexical sentiment

Option 2: the CLSTM model uses the auxiliary gradient descent algorithm to calculate the optimal solution for the Chinese vocabulary

Option 3: combined application of the CLSTM model in the auxiliary gradient descent algorithm and the emotional vocabulary formula is performed

Option 4: The comprehensive application of the CLSTM model superimposed on the SO-PMI algorithm, the auxiliary gradient algorithm, and the emotional vocabulary formula calculation is done at the same time

4.2. Comparison Results of Experimental Algorithms. In the above scheme, we carried out data comparison operations, respectively, and obtained the comparison results of the Chinese language sentiment analysis, as shown in Tables 1–4. Comparing the data in the table, we can clearly see that the data in scheme 1 is significantly lower than that of the other three schemes. We analyzed about 3000 words, but the algorithm's sentiment extraction and analysis efficiency is only 50.7%, which is far lower than other algorithms [17]. Scheme 2 adds an auxiliary gradient descent algorithm to the CLSTM model to process the optimal solution of the vocabulary, and the rate and accuracy of analyzing vocabulary sentiment are slightly improved, but the overall change is not large and there are still major problems. Scheme 3 then analyzes the changes brought by the emotion formula to the CLSTM model. It can be seen that there are obvious changes compared to Scheme 1, but the analysis of the emotional state after too many words is still unsatisfactory. After all the tests, we finally designed the fourth scheme and it was successfully selected. It has its uniqueness in addition to the advantages of other algorithms, which greatly improves the efficiency of sentiment analysis. Therefore, this scheme has obvious advantages compared with other schemes, and its computing speed and the accuracy of the language sentiment analysis make its comprehensive level the best [18].

We compared the line graphs of the analysis times for scenarios 1–4. The number of words from left to right in the figure is 100, 500, 1500, and 3000, respectively. We can clearly see the overall superiority of this scheme. When the number of vocabulary is larger, the more obvious the advantage is [19]. In this line chart, the larger the change of the line, the more obvious the distinction between the pros and cons of the model. We can clearly see from Figure 9 that when 100 text data is used as the standard, the line is basically a horizontal line with little change. When 3000 text data is used as the standard, the variation of schemes 1–4 is huge, which better verifies the superior performance of our selected algorithm.

The comparison effect is shown in Figure 10.

Here, we take the 3000 language vocabulary benchmarks as an example and list the line chart to compare the data of schemes 1–4 in detail. It can be clearly seen that the comprehensive performance of scheme 4 is significantly better than the other schemes [20]. The effect is shown in Figure 9.

4.3. Analysis of the SO-PMI Algorithm Model for Superposition Application. In the above case analysis, using this algorithm, when the number of the Chinese language vocabulary is about 2000, it can make the language sentiment analysis efficiency and language sentiment analysis accuracy achieve the best comprehensive performance.

Based on the comprehensive application of the CLSTM model superimposed on the SO-PMI algorithm proposed by the verification, that is, in the case of scheme 4, the number of 2000 Chinese language vocabulary is cited as the benchmark and the deep learning analyzes the emotional distribution of the vocabulary, conducts a comprehensive

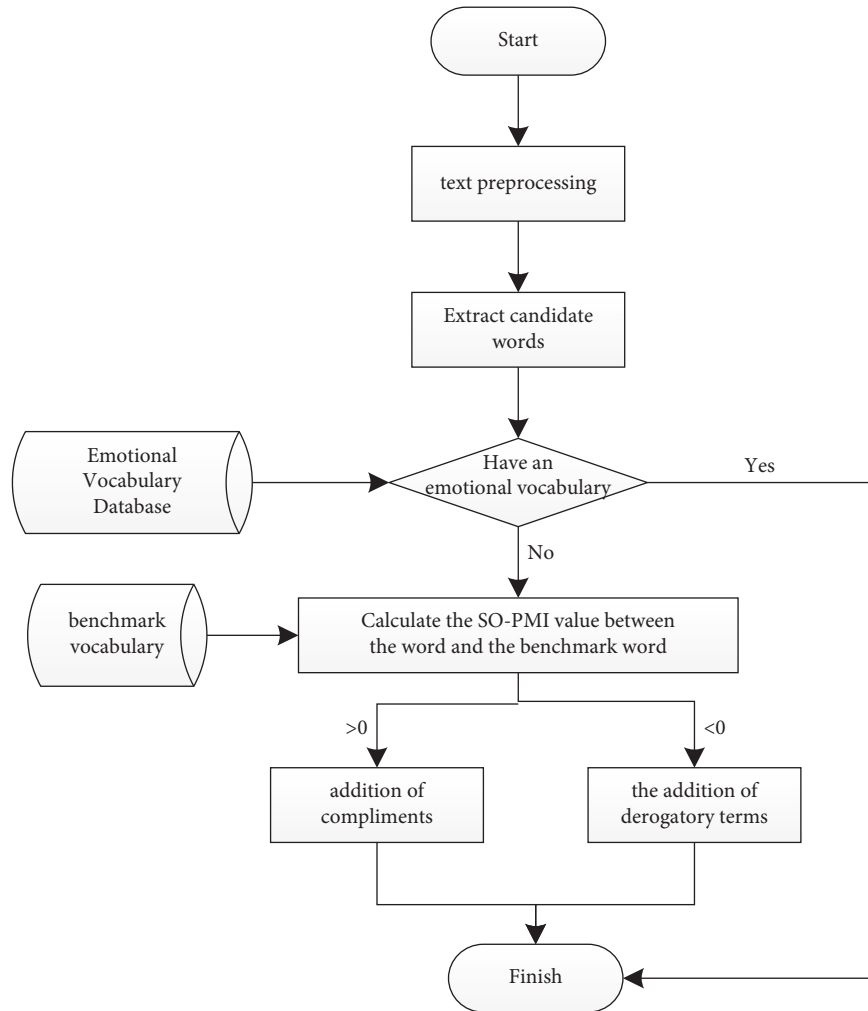


FIGURE 8: SO-PMI algorithm implementation.

TABLE 1: Chinese language sentiment extraction and analysis in scheme 1.

The number of emotion words extracted	Sentiment analysis rate (min)	Language sentiment analysis accuracy (%)
100	0.5	70.8
500	3.0	65.4
1500	10.1	59.6
3000	21.7	50.7

performance analysis on it, and shows the superior performance of the algorithm [21].

When the number of Chinese language vocabulary is 2000, its comprehensive performance index is shown in Figure 11.

When the number of words reaches 2000, it can meet the role of general article language sentiment analysis. The speed of language sentiment analysis is significantly improved compared with the previous algorithm. Based on 2000 words, the application waits about 4.5 minutes to meet people's general needs, which is the time to wait [22].

Text length is an important indicator of language sentiment analysis. If the text length is too short, it is difficult to

reflect the superior performance of each algorithm. If the text is too long, it will increase the burden of the algorithm model.

Therefore, we counted the distribution of language text length in general life, and Figure 12 is a graphic representation of its distribution. We control most of the text within a reasonable range (this article uses 3000 as the range), so the algorithm can effectively perform related sentiment analysis work [23]. The distribution of language texts in daily life is shown in Figure 12.

Similarly, we analyze the actual emotional needs in this field. The language emotion function can be applied to emotion retrieval, emotion summary, emotion question answering, and many places such as movie reviews, product

TABLE 2: Chinese language sentiment extraction and analysis in scheme 2.

The number of emotion words extracted	Sentiment analysis rate (min)	Language sentiment analysis accuracy (%)
100	0.4	78.9
500	2.8	70.5
1500	8.5	64.9
3000	16.4	60.5

TABLE 3: Chinese language sentiment extraction and analysis in scheme 3.

The number of emotion words extracted	Sentiment analysis rate (min)	Language sentiment analysis accuracy (%)
100	0.25	81.2
500	1.5	74.3
1500	6.0	65.9
3000	14.8	61.7

TABLE 4: Chinese language sentiment extraction and analysis in scheme 4.

The number of emotion words extracted	Sentiment analysis rate (min)	Language sentiment analysis accuracy (%)
100	0.11	98.8
500	0.6	89.3
1500	3.5	82.7
3000	8.1	77.6

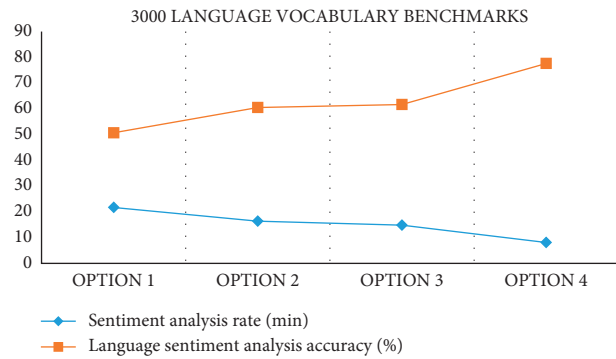


FIGURE 9: Comprehensive comparison of 4 algorithms on 3000 language vocabulary benchmarks.

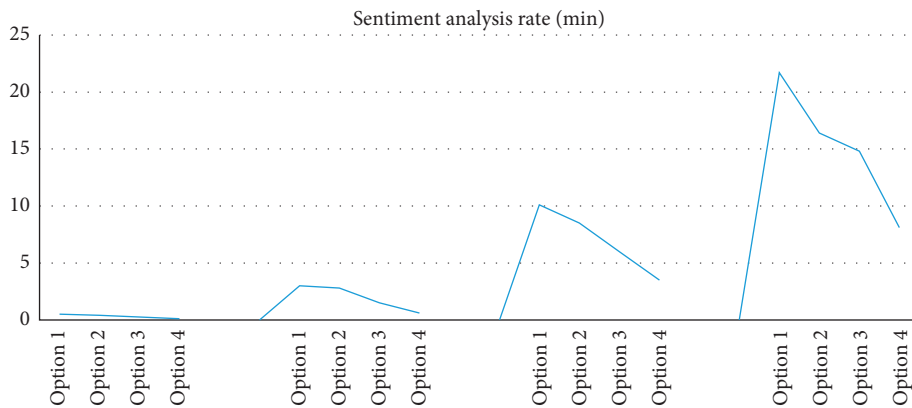


FIGURE 10: Overall comparison of sentiment analysis time for schemes 1–4.

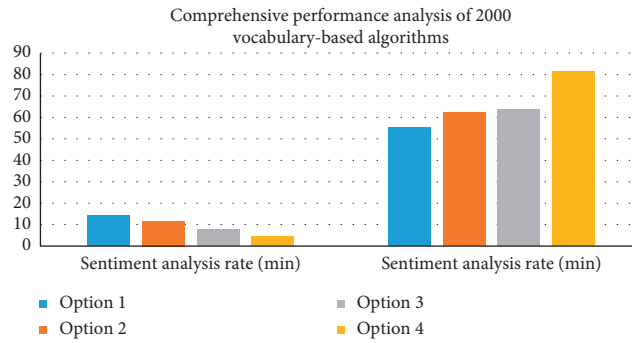


FIGURE 11: Comprehensive performance analysis of the 2000-word-based algorithm.

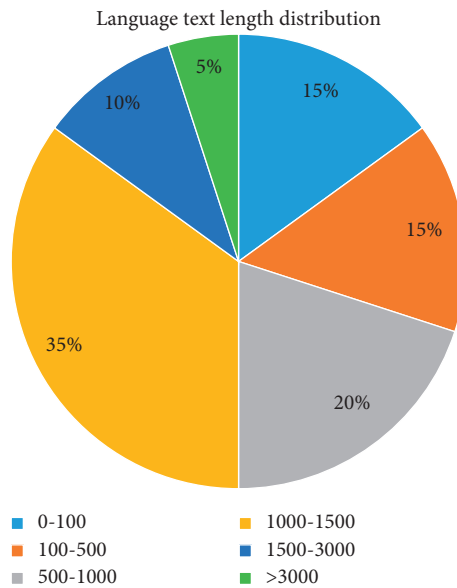


FIGURE 12: Statistical distribution display of language text length.

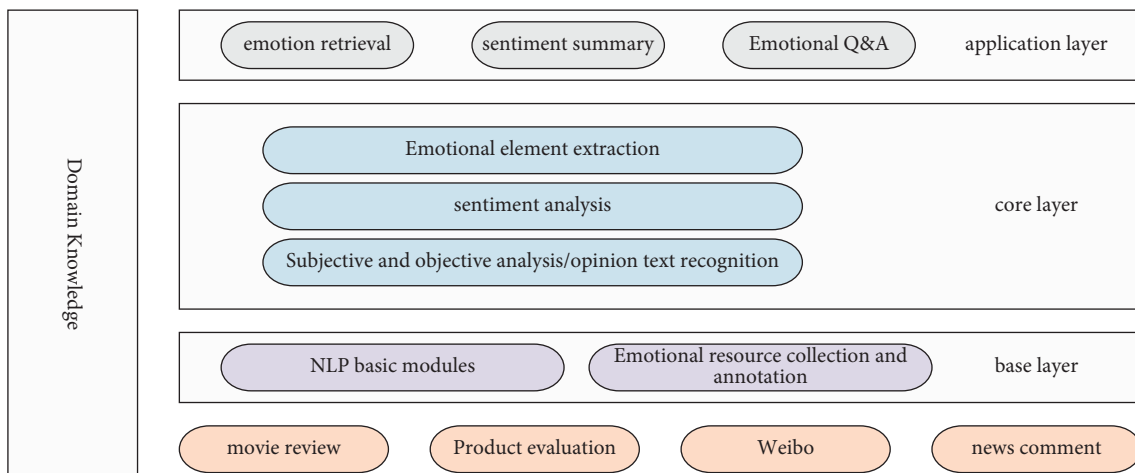


FIGURE 13: The main adaptation map of language emotion.

evaluations, microblogs, and news. It can be seen that its application field is broad, and the information given above shows that the comprehensive performance of the algorithm

is superior and it is easy to meet the needs of language sentiment analysis in people's daily life [24]. The overall block diagram of the language statistics is shown in Figure 13.

We need to transform traditional algorithmic sentiment analysis methods, improve the extraction and analysis of key words, and expand the coverage of key sentences to improve their representativeness. At the same time, emoji is also a major direction of its sentiment analysis. We also need to increase the related construction tasks of the symbol dictionary and combine it with text language analysis, expand the sentiment dictionary on the SO-PMI algorithm, and formulate a corresponding set of sentiment rules calculation.

5. Conclusion

As mentioned above, the deep learning of this research is used for Chinese language sentiment extraction and analysis, through the reference of the original CLSTM model, plus the calculation of the auxiliary gradient algorithm and the emotional vocabulary formula $f_sentiment(s_i)$ and the synthesis of the SO-PMI algorithm model. The application can substantially improve the efficiency and accuracy of language emotion extraction and analysis. The algorithm we propose performs de-redundancy operations on some unnecessary neutral words and can extract and analyze language sentiment after a series of word segmentation operations. Compared with other single algorithms, the comprehensive performance of this algorithm is the best, which significantly improves the efficiency and accuracy of sentiment analysis. Under the premise of 2000-word vocabulary, the algorithm fully meets the needs of daily life. In the traditional language sentiment analysis work, we add the new text extraction opinion vocabulary to explore the mutual emotional connection between vocabulary and objective affairs [25]. Combined with the extraction of opinion relations in the application of tree kernels, a new kernel reference is developed, multiple analyses is performed on the emotional vocabulary data model, and finally, the overall performance improvement of the efficiency of language sentiment analysis is achieved. The algorithm also has imperfections. Our emotional vocabulary is imperfect, and the Chinese text language is extremely difficult to analyze. In addition, there are some special sentence structure analyses, such as antonyms. We still need to increase the extraction of key sentence patterns and explore more advantageous algorithms to make the extraction of text information more representative.

Data Availability

The experimental datasets used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

References

- [1] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Journal of the American Medical Association*, vol. 316, no. 22, p. 2402, 2016.
- [2] G. Litjens, T. Kooi, B. E. Bejnordi, A. Setio, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.
- [3] B. C. Ng, C. Cui, and F. Cavallaro, "The annotated lexicon of Chinese emotion words," *Word*, vol. 65, no. 2, pp. 73–92, 2019.
- [4] Z.-T. Liu, M. Wu, W.-H. Cao, Y. Mei, and J.-W. Mao, "Speech emotion recognition based on an improved brain emotion learning model," *Neurocomputing*, vol. 309, pp. 145–156, 2018.
- [5] Y. Chen, Z. Lin, Z. Xing, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2017.
- [6] S. Ding and R. A. Saunders, "Talking up China: an analysis of China's rising cultural power and global promotion of the Chinese language," *East Asia*, vol. 23, no. 2, pp. 3–33, 2006.
- [7] A. Sengupta, S. Roy, and G. Ranjan, "LJST: a semi-supervised joint sentiment-topic model for short texts," *SN Computer Science*, vol. 2, no. 4, p. 256, 2021.
- [8] Y. H. Lee and K. S. Ang, "Brand name suggestiveness: a Chinese language perspective," *International Journal of Research in Marketing*, vol. 20, no. 4, pp. 323–335, 2003.
- [9] R. V. Kübler and A. Colicev, "Social media's impact on the consumer mindset: when to use which sentiment extraction tool?" *Journal of Interactive Marketing*, vol. 50, no. 1, pp. 136–155, 2020.
- [10] Y. Wu, W. Gang, and H. Li, "A Chinese word segmentation algorithm based on N-gram model and machine learning," *Journal of Electronics and Information*, vol. 11, pp. 1148–1153, 2001.
- [11] S. Wan, J. Tang, and X. Yang, "Research on big data of customer product reviews based on text sentiment extraction and statistical analysis," *Academic Journal of Humanities & Social Sciences*, vol. 3, no. 3, 2020.
- [12] V. Radhakrishnan, C. Joseph, and K. Chandrasekaran, "Sentiment extraction from naturalistic video," *Procedia Computer Science*, vol. 143, pp. 626–634, 2018.
- [13] M. Tubishat, N. Idris, M. A. M. Abushariah, and M. Abushariah, "Implicit aspect extraction in sentiment analysis: review, taxonomy, oppportunities, and open challenges," *Information Processing & Management*, vol. 54, no. 4, pp. 545–563, 2018.
- [14] P. G. Preethi, V. Uma, and A. Kumar, "Temporal sentiment analysis and causal rules extraction from tweets for event prediction," *Procedia Computer Science*, vol. 48, pp. 84–89, 2015.
- [15] R. Shettar and S. Taorem, "Sentiment extraction and analysis of product reviews at sentence level," *Current Trends in Information Technology*, vol. 1, 2011.
- [16] S. Wang, A. Yang, and D. Li, "Research on sentence sentiment tendency classification based on Chinese sentiment vocabulary," *Computer Engineering and Applications*, vol. 45, no. 24, pp. 153–155+161, 2009.
- [17] Y. Zhang and H. Wang, "Chinese text sentiment analysis based on att-BiGRU-CRF model," *Journal of Tianjin University of Technology*, vol. 37, no. 6, pp. 31–35, 2021.
- [18] Z. Chen, Y. Qian, and W. Zhao, "Research on sentiment classification model of bullet screen text—based on Chinese pre-training model and bidirectional long short-term memory network," *Journal of Hubei University of Technology*, vol. 36, no. 6, pp. 56–61, 2021.

- [19] Z. Pan, L. Zhao, L. Yuan, and H. Wang, "FastText Chinese sentiment polarity analysis based on Borderline-Smote algorithm improvement," *Computer Applications and Software*, vol. 38, no. 11, pp. 295–299+349, 2021.
- [20] B. Zhang, H. Zhang, T. Li, and J. Shang, "A sentiment analysis method for Chinese reviews based on multi-input model and syntactic structure," *Big Data*, vol. 7, no. 6, pp. 41–52, 2021.
- [21] T. Diao, J. Zhang, C. Yao, and W. Li, "Application of Chinese text sentiment classification based on recurrent neural network," *Wireless Internet Technology*, vol. 18, no. 19, pp. 96–97, 2021.
- [22] Y. Yuan, "Research on sentiment classification of online review texts based on Naive Bayes," *Inner Mongolia Science and Technology and Economy*, vol. 18, pp. 91–94, 2021.
- [23] H. Zhang, H. Huang, and W. Li, "Speech emotion database for emotion change detection," *Computer Simulation*, vol. 38, no. 9, pp. 448–455, 2021.
- [24] Y. Wang, "On the emotional penetration and integration in the process of Chinese teaching in higher vocational colleges," *University*, vol. 35, pp. 128–130, 2021.
- [25] Z. Huang, X. Wu, Y. Wu, and J. Ling, "Chinese text sentiment classification combined with BERT and BiSRU-AT," *Computer Engineering and Science*, vol. 43, no. 9, pp. 1668–1675, 2021.