Sentiment Analysis of User Comment Text based on LSTM

FENG LI, CHENXI CUI, YASHI HU, LINGLING WANG*
School of Management Science and Engineering,
Anhui University of Finance and Economics,
Bengbu 233030,
CHINA
Corresponding Author

Abstract: Taking the user-generated Chinese comment dataset on online platforms as the research object, we constructed word2vec word vectors using gensim and built a sentiment analysis model based on LSTM using the TensorFlow deep learning framework. From the perspective of mining user comment data on the platform, we analyzed the sentiment tendency of user comments, providing data support for hotels to understand consumers' real sentiment tendencies and improve their own service quality. Through analysis of the validation dataset results obtained by crawling the website, the accuracy of this LSTM model can reach up to 0.89, but there is still much room for improvement in the accuracy of sentiment analysis for some datasets. In future research, this model needs further optimization to obtain a stable and more accurate deep-learning model.

Key-Words: Online; Comment Text; LSTM; Sentiment Analysis.

Received: May 8, 2022. Revised: February 12, 2023. Accepted: March 3, 2023. Published: April 5, 2023.

1 Introduction

In recent years, with the rapid development of Internet technology and the arrival of the era of universal 5G, Internet applications and big data computing are involved in every aspect of daily life. By the end of 2021, the global Internet user base has reached 4.9 billion. Compared with 2019 (401 billion), the global Internet user case has grown by 19.5%, with an increase of 800 million new Internet users. Among them, the growth rate of global Internet users reached 10.2% in 2020, the highest in a decade. As a result, the Internet generates around 4 PB of data every day, including about 10 billion text messages. With the increasingly widespread use of online shopping, the amount of information data generated by users of e-commerce platforms is also increasing, with a considerable portion being emotional evaluations, opinions, and thoughts. Extracting user sentiment from such a large and complex volume of textual information and identifying users' sentiment tendencies has become a research area of great interest in the field of natural language processing. Sentiment analysis, also known as opinion mining, is the process of classifying text into positive, negative, and neutral sentiments, [1]. The main purpose of sentiment analysis is to classify text into positive, negative, and neutral sentiments. To date, sentiment analysis has been extensively explored and breakthroughs have been made in various research methods. Common methods for sentiment analysis include traditional sentiment lexicons and machine learning. The traditional sentiment lexicon method involves comparing the emotional tendency and information intensity of the vocabulary in the pre-constructed sentiment lexicon with the content of the text and then classifying the text. Traditional sentiment lexicon methods can use existing high-quality lexicons, but the lexicon cannot cover all the vocabulary in all fields, especially with the impact of the newly emerged internet language. Machine learning models text classification by modeling certain features and learning from manually labeled data, but this method requires some manual feature engineering.

The content of user comments contains the subjective emotions of the user. Sentiment analysis of user message content usually refers to the mining, parsing, generalisation, and inference of user-generated content based on machine learning, from which user attitudes and opinions are obtained. Sentiment analysis can be divided into chapter-level sentiment analysis, sentence-level sentiment analysis, and aspect-level sentiment analysis, with chapter-level sentiment analysis and sentence-level sentiment analysis belonging to coarse-grained sentiment analysis and aspect-level sentiment analysis belonging to fine-grained sentiment analysis. Coarse-grained sentiment analysis can only analyse the sentiment of a whole text or paragraph, but not the multiple perspectives contained in a large text. In contrast, fine-grained sentiment analysis can determine the sentiment of individual entities more accurately, making aspect-level sentiment analysis a hot research area in the field of sentiment analysis. Aspect-Based Sentiment Analysis (ABSA) is a fundamental task in sentiment analysis, which aims to identify the aspects present in a sentence and determine the sentiment polarity of each aspect, 0.

Through research on relevant domestic and foreign work, we found that the mainstream text sentiment analysis methods currently mainly include: sentiment analysis based on sentiment dictionaries, sentiment analysis based on machine learning, and sentiment analysis based on deep learning. Sentiment analysis based on sentiment dictionaries and rules first compares the text content with specific entities in the sentiment dictionary to derive sentiment values, and then the results are weighted to derive the sentiment tendency of the text. This method can produce good sentiment analysis results if the sentiment lexicon is sufficiently rich. A machine learning algorithm is a generic term for a class of algorithms that construct a function with a large amount of data as input, the output of which can be classification, prediction, etc. And this function is equally applicable to new sample data. Machine learning-based sentiment analysis is a method of constructing a function model into which a large amount of textual information is input and extracting features through machine learning algorithms. KNN, NB, and SVM are common algorithms used in machine learning. Deep learning is a practical application of multilayer neural networks in learning, which is still essentially in the realm of machine learning, although it can solve complex problems that are difficult to solve with traditional machine learning, so it has been singled out as a separate discipline. Common models of deep learning include CNN, RNN, Transformer, GRU, and LSTM. The underlying structure of a neural network is shown in Figure 1 and contains an input layer, a hidden layer, and an output layer. Each neuron in the input layer can be used as a feature of an object, the hidden layer may have multiple layers and it will transform the information from the input into something that can be used in the output layer, and the output layer transforms the results of the hidden layer into the desired result.

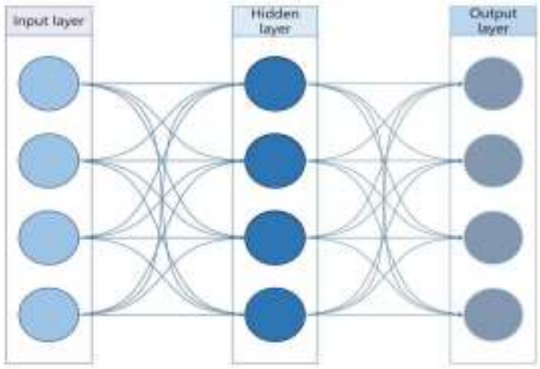

Fig. 1: Neural network structure

2 Related Works

2.1 Studies Related to Sentiment Analysis

2.1.1 Sentiment Analysis Methods based on Sentiment Dictionaries

The earliest English sentiment lexicon that appeared is SentiWordNet, in addition to the commonly used ones, such as General Inquirer, Opinion Lexicon, and MPQA. Chinese emotion dictionaries are widely used, such as HowNet and Dalian University of Technology's Chinese emotion vocabulary ontology database. In addition, 0, proposed a sentiment lexicon for Vietnamese, which includes more than 100,000 Vietnamese emotional vocabularies. The slang sentiment dictionary SlangSD, which was built by [4], from web resources, is also effective in identifying the sentiment of users. [5], used a constructed dictionary of network words, degree adverbial dictionaries, negation dictionaries, and other related dictionaries, and trained them with the help of Weibo texts to derive updated sentiment values. [6], proposed a sentiment classifier that trains incremental words from time-varying distributional word vectors, automatically extracts continuously updated sentiment vocabulary from Twitter streams, and obtains a time-varying sentiment lexicon based on incremental word vectors. [7], integrated emojis, modifiers, and domain-specific terms to analyse comments posted by online communities, overcoming the limitations of previous methods. Compared to the general approach, sentiment analysis was greatly improved by integrating modifiers, emojis, negation words, and domain-specific terms. However, the sentiment analysis method based on sentiment dictionaries relies too much on sentiment dictionaries, and in today's information age, a single sentiment dictionary cannot make accurate judgments, while building a more complete and diverse dictionary can be labor-intensive.

2.1.2 Machine Learning-based Sentiment Analysis Methods

[8], compared the results of decision trees, Bernoulli NB (BNB), Maximum Entropy (ME), support vector machines (SVM), and multinomial naive Bayes (MNB) in sentiment classification, and found that multinomial naive Bayes obtaining the best results of 88. 5%. [9], constructed a sentiment analyzer based on SVM and naive Bayes to analyze Twitter data and compared it with a sentiment analyzer using only SVM or NB. [10], proposed an optimized sentiment analysis framework (OSAF), which uses SVM lattice search techniques and cross-validation. [11], proposed an emoticon-based sentiment analysis method and discussed the role of symbolic expressions in sentiment analysis. [12], proposed a computational algorithm for semantic analysis based on the WordNet linguistic English lexicon training set, using a combination of machine learning algorithms SVM and NB to automatically detect strongly associated negative tweets.

Although machine learning-based sentiment analysis has made progress compared to lexicon-based sentiment analysis, it still requires manual labeling of text and subjective factors can affect the final result. Traditional machine learning requires high model requirements, and if the model is not efficient, it is difficult to adapt to the era of exploding information. In addition, traditional machine learning has difficulty using contextual information in sentiment analysis, which also affects accuracy.

2.1.3 Deep Learning-based Sentiment Analysis Methods

A sentiment analysis method based on deep learning can automatically learn deep features from a large amount of text information, and the sentiment analysis is effective and the model is highly adaptable without human intervention during the learning process. [13], proposed a Restricted Boltzmann Machine (RBM) based rule model for sentiment analysis of sentences. [14], proposed a restricted data framework using RNN as a framework to train a single model using the largest dataset of languages and reuse it for languages with limited datasets. This framework has good results for sentiment analysis of small languages. LSTM is a special structure of RNN, and to improve the training speed and reduce computational cost and time, [15], proposed an attention-based LSTM oriented aspect-level sentiment memory network classification for sentiment classification based on LSTM, [16], proposed a streamlined LSTM with six different parameters and compared the performance differences between these LSTMs using the Twitter dataset to establish the best set of parameters for the LSTM, [17], proposed a new sentiment analysis scheme based on Twitter and Weibo data, focusing on the impact of expressions on sentiment, and training an emotion classifier by attending to these binary expressions, embedded in an attention-based long- and short-term memory network, which is a good guide for sentiment analysis. Because of the lower human input as well as the higher accuracy, deep learning-based sentiment analysis methods have become a hot research topic in recent years.

2.1.4 Analysis of Irony

It is easy to find that there are a lot of phenomena of irony and sarcasm on online platforms, and the emotion implied by such statements is often the opposite of the surface meaning of the statements. Therefore, the analysis of ironic statements and the analysis of the deeper meaning of the statements will help to determine the emotional polarity of the text. [18], achieved good results in experiments with four machine learning methods by improving the sentiment analysis process and decision-making process and crawling data on Twitter, linear SVC (accuracy=83%, f1-score=0.81), logistic regression (accuracy=83%, f1-score=0.81), Naïve Bayes (accuracy=74%, f1- score=0.73) and random forest classifier (accuracy=80%, f1-score=0.81). Some authors, [19], found that previous research on sarcasm detection has mostly been conducted using natural language processing techniques, without considering the context, user's expression habits, etc. Therefore, a two-channel convolutional neural network was used to analyze the semantics of the target text, as well as its emotional context, and to extract the user's expression habits using an attention mechanism. The effectiveness of the method is confirmed by experiments on several datasets, and it can effectively improve the performance of the irony detection task.

2.1.5 Implicit Sentiment Analysis

Implicit sentiment analysis is a special part of the sentiment analysis field because of the lack of sentiment vocabulary and the ambiguity of sentiment polarity, which is a difficult area of research at this stage. Combing the literature on implicit sentiment analysis at this stage, it is found that the current research is very limited. [20], found that previous Graph Convolutional Networks (GCNs) used for the study of sentiment analysis problems had difficulty in effectively using contextual context or often ignored the dependencies between phrases. Therefore, they proposed a context-specific heterogeneous graph convolutional network (CsHGCN) based on this, and experimental results showed that the model could effectively identify target emotions in sentences.

2.1.6 Aspect-level Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA), an actively challenging part of the sentiment analysis field, aims to identify and analyze the fine-grained sentiment polarities towards particular aspects.

[21], proposed a new neural network-based framework to analyze the sentiment of aspect targets in comments. This framework captures distant textual sentiment information through a multi-attentive mechanism, employing a non-linear combination with recurrent neural networks to enhance the expressive power of the model, allowing it to handle more complex semantic problems. The performance of this model is also validated on four datasets (two from SemEval2014 (restaurant and laptop reviews), a Chinese news review dataset, and Twitter datasets).

[22], found that most previous prediction methods used long- and short-term memory and attention mechanisms to analyze the emotional polarity of the target of interest, and that such methods tended to be more complex and required more training time. Therefore, it was proposed to group the previous methods into two subtasks: aspect-category sentiment analysis (ACSA) and aspect-item sentiment analysis (ATSA). A model based on gating mechanisms and convolutional neural networks is also proposed, which is more accurate and effective. The method firstly uses a new gating unit, Tanh-ReLU, to selectively output sentiment features based on a given entity or aspect; this architecture is simpler than the attention layer used in existing models; secondly, the computations of this model are easily deserialized during training and the gating unit works independently, and finally, experiments on the SemEval dataset validate the effectiveness of the model.

Arabic poses several challenges for the task of sentiment analysis in Arabic because of its complex grammatical structure and the lack of relevant resources. Some scholars have taken the sentiment analysis of aspects of Arabic as a research direction, [23], and used a composite model combining a long short-term memory (LSTM) model and a convolutional neural network (CNN) to analyze the sentiment of Arabic tweets. For the Arabic sentiment tweet dataset (ASTD), this model scored 64.46% on F1, outperforming other deep learning models; some scholars, [24], research using two different long short-term memory (LSTM) neural networks for aspect-level sentiment analysis of Arabic hotel reviews. The first is an aspect-OTEs oriented LSTM for aspect sentiment polarity classification as sentiment polarity markers, and the second is a character-level bidirectional LSTM along with a conditional random field classifier (Bi-LSTM-CRF) for aspect opinion target expression (ballot) extraction. This method was evaluated using a reference dataset of Arabian hotel reviews and the results showed that this method outperformed the baseline study on both tasks by 6% and 39% respectively.

2.2 Relevant Research Techniques

Sentiment analysis of text content is the complete process of text preprocessing such as word segmentation, stop-word removal, and named entity recognition on the target text, followed by text vectorization, feature engineering, model training, classifier, and other processes to derive sentiment tendency labels. A flowchart of text classification is presented in Figure 2.

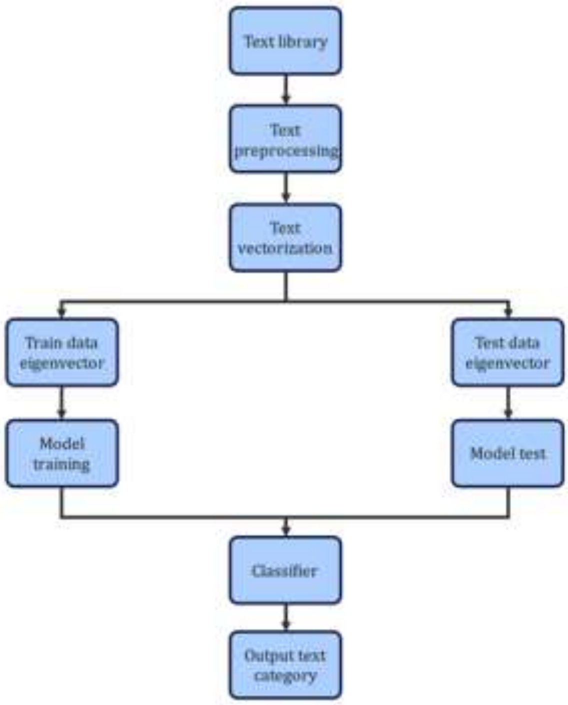

Fig. 2: Flowchart of text classification

2.2.1 Text Pre-Processing

The implementation of text preprocessing mainly involves word segmentation of txt format texts. This experiment uses the jieba Chinese word splitting tool, a widely used and effective Chinese word splitter with exact mode, full mode, and search engine mode that allows for precise segmentation of text sentences, fast scanning of the entire content, and secondary recall segmentation of long words. It is also possible to improve the segmentation effect by manually defining proper nouns in the text. In practice, there are a large number of intonational auxiliaries, personal pronouns, and other words that are not related to emotional tendencies, which can be filtered by building a deactivation dictionary. In addition, there are different dictionaries for different domains, [25]. The dictionaries are designed to include new words and specialized words that are unique to the field.

2.2.2 Text Vectorization

Text vectorization refers to converting Chinese text content that cannot be recognized by a computer into a vector form with digitized features that can be recognized by a computer. In this paper, word2vec is used to complete text vectorization, transforming the text preprocessed dataset into a vector with uniform dimensions, thus completing the simplification of shifting data from high latitude to low latitude. Depending on the definition of output and input, two algorithms can be classified: Skip-gram and CBOW. The CBOW algorithm is a three-layer neural network that predicts target words from contextual words, defining the words of the context in which a word is located as input and itself as output, using a corpus of corpora for training, and calculating the vector values of the context in the projection layer and summing them to output information about the target word. The Skip-gram algorithm reverses the causality of CBOW by defining the words in the context of the target word as the output, and the words themselves as the input, predicting the information of the contextual words with the help of the target word.

2.2.3 Recurrent Neural Network (RNN)

Recurrent Neural Networks are a class of sequences data as input and perform in the direction of evolution of the sequence recursively and all nodes (recurrent units) are connected in a chain-like manner in recurrent neural networks. Recurrent Neural Networks have memory, share parameters, and are Turing-complete and are therefore very useful in the analysis of sequential non-linear features. Recurrent neural networks are used in natural language processing such as speech recognition, language modelling, and machine translation, and also for various time-series predictions. The introduction of convolutional neural networks constructed recurrent neural networks that can process sequential inputs containing computer vision problems.

2.2.4 Long and Short-Term Memory

Long short-term memory (LSTM) is a special kind of RNN that was designed to address the problem of gradient vanishing and explosion during the training of long sequences. LSTM is a variant of RNN, with the core concept of cell states and "gate" structure. Cell states are the equivalent of information transmission paths that allow information to be passed along in a sequence. You can think of it as the 'memory' of the network. Theoretically, the cell states can pass on information relevant to the sequence processing all the way through. Thus, even information from earlier time steps can be carried to cells at later time steps, which overcomes the effects of short-term memory. Information is added and removed by means of 'gate' structures, which are trained to learn which information to keep or forget. In recent years, recurrent neural networks have been used in speech recognition, image processing, ECG arrhythmia classification, and natural language processing, and so on [26]. The LSTM model is shown in Figure 3.

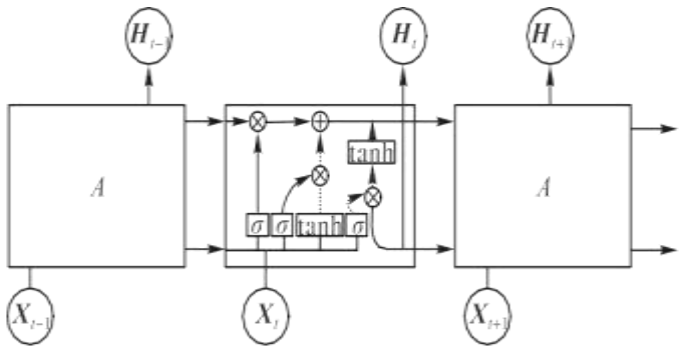

Fig. 3: LSTM structure

3 Algorithm

Due to the over-reliance on sentiment lexicons, simple lexicons cannot accurately discriminate sentiments, while constructing a more complex and diverse lexicon will require a large amount of manpower and resources. Although sentiment analysis based on machine learning has made progress compared to lexicon-based sentiment analysis, subjective factors have a significant impact during manual calibration. Moreover, traditional machine learning models require high model accuracy, and the explosion of information in today's world makes it difficult for models to adapt perfectly to complex and varied needs. In addition, traditional machine learning has difficulty utilizing contextual information, which can affect accuracy in sentiment analysis. Deep learning-based sentiment analysis methods can automatically learn deep features from a large amount of text information, with good sentiment analysis effects and strong model adaptability, without the need for human intervention in the learning process. Due to the low efficiency and quality of traditional methods, people have begun to use deep learning to construct network models for text classification tasks. [30], reviewed more than 150 deep learning-based text classification models developed in recent years in their review and discussed their technical contributions, similarities, and advantages. Therefore, this paper chose a deep learning-based sentiment analysis method to complete the sentiment judgment of text information.

Common deep-learning models include CNN, RNN, Transformer, GRU, and LSTM. Traditional CNN models may not activate neurons that recognize the same object slightly differently due to translational invariance, i.e., changes in the orientation or position of the same object. Moreover, the pooling layer causes a significant loss of valuable information, ignoring the correlation between local and global features. Therefore, CNN models are difficult to accurately judge the precise textual sentiment. Although RNN models can consider historical information during calculation and share weights over time compared to CNN models, their computation speed is slow and cannot consider any future input of the current state. In addition, RNN models often suffer from gradient disappearance and explosion because it is difficult to capture long-term dependencies, and multiplication gradients can decrease or increase exponentially with the number of layers. Although GRU models can effectively alleviate the problem of gradient explosion in RNN models, compared to GRU models, LSTM models have more parameters, stronger functionality, and stronger expressive power.

LSTM has a similar working mechanism to RNN, but its implementation of more refined internal processing units enables effective storage and updating of contextual information. Due to its excellent properties, LSTM has been used in many tasks related to sequence learning, such as speech recognition, [31], language models, [32], part-of-speech tagging, [35], and machine translation, [36]. Therefore, considering all factors, this paper uses LSTM as the deep learning model for sentiment analysis.

3.1 Recurrent Network Model

RNNs, or Recurrent Neural Networks, excel in processing sequences of data where context is essential. One of the distinguishing features of RNNs is their ability to create directed loops between nodes, [38]. Examples of sequence data that RNNs can handle well include speech recognition, language prediction, garbage image classification, [39], and stock data analysis, [40]. Since the data at each node in the sequence is related to the preceding and subsequent data points, RNNs can capture these dynamic relationships. By retaining previous information and using it as input for subsequent nodes, RNNs are ideal for analyzing time-sequenced data.

3.2 RNN Model Gradient Disappearance Phenomenon

[41], proposed that standard RNNs suffer from gradient vanishing, which refers to the vanishing of gradients in RNNs for more distant time steps. The BPTT method is used for backpropagation in RNNs, where the gradient of loss against parameter W is equal to the sum of the derivatives of loss against W at each time step. This can be expressed mathematically as a formula.:

$$\frac{\partial E}{\partial W^{\square}}=\sum_{i=1}^{t}\frac{\partial E}{\partial y^{t}}\frac{\partial y_{t}}{\partial \square_{t}}\frac{\partial \square_{t}}{\partial \square^{i}}\frac{\partial \square_{i}}{\partial W^{\square}}$$

The calculation in the above equation is more complex and is based on a continuous derivative of the complex function.

$$\frac{\partial \square_{t}}{\partial \square^{i}}=\prod_{k=i+1}^{t}\frac{\partial \square_{k}}{\partial \square^{k-1}}$$

$\frac{\partial \square_{k}}{\partial \square^{k-1}}$ is the partial derivative of the current hidden state with respect to the previous hidden state.

$$\frac{\partial \square_{k}}{\partial \square^{k-1}}=\sigma' W^{\square}$$

Suppose that a time step j is (t-j) moments away from time step t. So:

$$\frac{\partial \square_t}{\partial \square^i} = \prod^{t-j} \sigma' W^{\square}$$

If t-j is large, that is, j is far from the t time step, when$\sigma' W^h$>1, a gradient explosion problem arises and $\sigma' W^h$ <1, there is a gradient disappearance problem. And when t-j is small, there is no gradient disappearance/gradient explosion problem. In summary, the gradient of j farther away from time step t will vanish and j does not affect the final output y^t has no effect on the final output. This means that there can be no long-term dependence on RNN.

3.3 Gradient Disappearance Phenomenon

To address the problem of long-term dependencies, [42], proposed a Long Short-Term Memory (LSTM) network, which performs much better than RNN, especially in long-distance dependent tasks, [43]. The LSTM was originally designed so that the bias derivative of the current memory unit with respect to the previous memory unit would be constant. As in the original version of the LSTM in 1997, the memory cell update formula was

$$C^t = C^{t-1} + Z^i \odot x^t$$

$$\frac{\partial C_t}{\partial C^{t-1}} = 1$$

Later, to avoid the wireless growth of memory cells, **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**, later refined the LSTM cell by introducing the "forget gate". The updated formula is:

$$C^t = Z^f \odot C^{t-1} + Z^i \odot x^t$$

The value of the partial derivatives at this moment is:

$$\frac{\partial C_t}{\partial C^{t-1}} = Z^f$$

AlthoughZ^f is a value in the interval [0,1], not in the sense of satisfying the bias of the current memory cell to the previous memory cell as a constant. However, it is common to set a large bias term to the forgetting gate such that the forgetting gate is closed in most cases and open only in a few cases. Recall the formula for the forgetting gate, here we have added the bias b.

$$Z^f = \sigma(W_f[\square^{t-1}, x_t] + b^f)$$

The forget gate is closed when it tends to 1 and opened when it tends to 0. By setting a large bias term, most forget gates tend to 1. By setting a large bias term, most of the forgetting gates converge to 1. This also alleviates the problem of gradient disappearance due to fractional multiplication.

4 Sentiment Analysis based on LSTM

4.1 Data and Processing

In this paper, the dataset is based on the comment corpus collated by Tan Songbo, with 2000 positive and negative examples each, which is a relatively small dataset. Examples are shown in Table 1. Moreover, Table 2 presents model parameter settings.

Table 1. Example of ChnSentiCorp data

Positive	Negative
It is a very nice 5-star hotel, the rooms are large, the facilities are new, and the location is convenient to the financial center, so I would consider staying there again.	Depressed!!! Angry!!! I don't understand that the fiber optic is even slower than the internet speed in Shanghai Jinjiang Star, don't go to this place if you want fast internet speed at night!!!!
The room was clean, the facilities were ok, the furniture was a bit old. The business room has a good floor front desk and the price point is relatively low for a 4-star.	The room was never arranged to have a frontal lake view, especially as the standard of the reception was really poor, with grumbling and expressionless faces.
The hotel was clean, the waiter would recommend me to the ladies' non-smoking floor, the facilities were better, and the dim sum in the restaurant tasted ok.	Not as bad as a good 2-star or no-star hotel

Table 2. Model parameter settings

Dim	Words	Buffer	LSTM_size	Dropout	Epochs	Batch size
300	300	3500	32	0.5	25	20

Word vectors: This experiment uses open-source word vectors and Chinese-word-vectors The Word Vector is a Word Vector trained from the Zhihu corpus.

In this work, the data was divided into a training set and a test set in a ratio of 4:1. For the training and validation sets, the following format was followed when producing the training data: In the text file, each row is the input for one sample, where each paragraph is commented on for one line and separated from the word by space using jieba.

4.2 Measurement Criteria

In this paper, recall, accuracy, precision, and F1 values are used as experimental measures and positive texts are used to refer to texts with positive affective tendencies and negative texts to refer to texts with negative affective tendencies. In the above confusion matrix, TP is the number of texts correctly classified as positive; FN is the number of texts incorrectly classified as positive; FP is the number of texts incorrectly classified as negative; and TN is the number of texts correctly classified as negative.

Precision is the percentage of texts judged to be of a certain type that is correctly judged.

$$p = \frac{TP}{TP + FP} \text{ or } \frac{TN}{TN + FN}$$

The recall is the percentage of texts that are actually of a certain type that are judged to be correct.

$$R = \frac{TP}{TP + FN} \text{ or } \frac{TN}{TN + FP}$$

The F1 value is the summed mean value of precision and recall, which corresponds to the combined precision and recall evaluation metric.

$$F1 = \frac{2 * P * R}{P + R}$$

Accuracy is the percentage of correctly judged texts out of all texts.

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

4.3 Parameter Settings

The experiment in this paper used the open-source word embedding model from Zhihu to train text information into 300-dimensional word vectors. The parameters of the LSTM model were set as follows: the maximum word count was set to 300 (setting the dimension too high would result in longer training time); a buffer zone of 3500 was reserved; the regularization parameter was set to 0.5; the batch size was set to 20; and the algorithm worked 25 times on the entire training dataset.

4.4 Results

Table 3. Experiment results using LSTM

Textual Emotional Tendencies	Results		
	Positive	Negative	support
Positive texts	TP: 865	FN: 135	1000
Negative text	FP: 91	TN: 909	1000

Table 4. LSTM model processing data results

	precision	recall	f1-score	support
POS	0.90	0.87	0.88	1000
NEG	0.87	0.91	0.89	1000
micro avg	0.89	0.89	0.89	2000
macro avg	0.89	0.89	0.89	2000
weighted avg	0.89	0.89	0.89	2000

4.4.1 Model Training Results

In this paper, 5000 positive and 5000 negative emotion texts were used for the training of the model, which were divided into a training set and a test set according to 4:1, with 8000 texts in the training set, 4000 positive and 4000 negative emotions texts in the training set, and 2000 texts in the test set, containing positive and the test set contains 2000 texts, including 1000 positive and 1000 negative texts. After processing the LSTM model, the following results were obtained.

In Table 3 we present the experiment results using LSTM. Similarly, in Table 4 we present the LSTM model processing the data results of our paper. Specifically, regarding Table 4 properties, we specify the following:

1. Macro average macro avg: sums the accuracy, recall, and F1 values for each category to find the average.
2. Micro avg builds a global confusion matrix for each instance in the dataset, regardless of category, and then calculates the corresponding metric.
3. weighted avg: an improvement on macro-averaging, considering the number of samples in each category as a proportion of the total sample.

4.4.2 Validation of the Dataset Results

By importing a corpus of e-commerce reviews from Baidu's library into the trained model, containing 1000 positive and negative texts each, the 2000 texts were divided into 20 groups of data, and the accuracy, recall, F1 value, and accuracy of these 20 groups were calculated. The following graphs were generated from the results.

According to the analysis of the above graphs, we can find that: the accuracy of the positive text can reach a maximum of 0.98 and a minimum value of 0.82; the accuracy of the negative text can reach a maximum of 0.92 and a minimum value of 0.44; the recall of the positive text can reach a maximum of 0.91 and a minimum value of 0.64; the recall of the negative text can reach a maximum of 0.98 and a minimum value of 0.68; the accuracy of the positive text The maximum F1 value for positive text is 0.8727 and the minimum value is 0.7189; the maximum F1 value for negative text is 0.8383 and the minimum value is 0.5626; the accuracy of this LSTM model can reach up to 0.89.

A comprehensive analysis of this LSTM model leads to the conclusion that the accuracy of this LSTM model still needs to be improved and further improvements are needed to achieve more accurate sentiment propensity analysis.

5 Conclusion

In this paper, the sentiment tendency analysis of e-commerce platform reviews is carried out by the LSTM model, which is trained and validated by an open dataset downloaded from the web. Our research findings are summarized in Figure 4 regarding our experimental results for the LSTM model validation dataset. Moreover, Figure 5 showcases the accuracy, recall, and F1 values for forward text whereas Figure 6 is for negative text. Lastly, Figure 7 presents the overall accuracy of the studied data sets.

The results show that the classification accuracy of this LSTM model can reach a maximum of 0.89, but there is still much room for improvement. The LSTM model implemented in this paper aims to judge the sentiment tendency of user-generated reviews on e-commerce platforms, to perform sentiment analysis on reviews on e-commerce platforms, and to provide a proven method for e-commerce platforms to judge the sentiment polarity of user reviews and extract keywords in the process of investigating user feedback, to provide data support for merchants to understand consumers' needs and real reviews, and to improve service quality in a targeted manner. It provides data support. Sentiment analysis of user reviews can effectively find out whether users identify with a shop, observe how much they like the product, help the management of the e-commerce platform to discover the strengths and weaknesses of the shop, improve the level of service and enhance user satisfaction.

The collective amount of data taken in this experiment is not large enough for effective analysis of non-semantic symbols and expressions, the model training takes too long, and there are individually large differences in the process of analyzing the accuracy of the validation set. The analysis of emoji information, the use of multiple parameters, and the optimization of the model will be the next research directions in the future. In subsequent research, a comparison between the optimized LSTM model and other neural network deep learning models will also be obtained with the increasing capability of text information recognition and generalization.

Based on the content of this paper, future research can be conducted in four areas. First, it can further optimize the sentiment analysis model and try to use more efficient and accurate deep learning models, such as pre-trained language models such as BERT and GPT, and combine with attention mechanisms to improve the model's performance. Second, it is necessary to explore how to deal with the challenges of semantic complexity and ambiguity in Chinese sentiment analysis, further improving the accuracy and robustness of the model. Finally, it is necessary to consider the evolution of emotions and contextual factors to more accurately determine the user's emotional tendencies. These research directions will help further improve the effectiveness of text sentiment analysis based on LSTM and make it more applicable to practical scenarios.

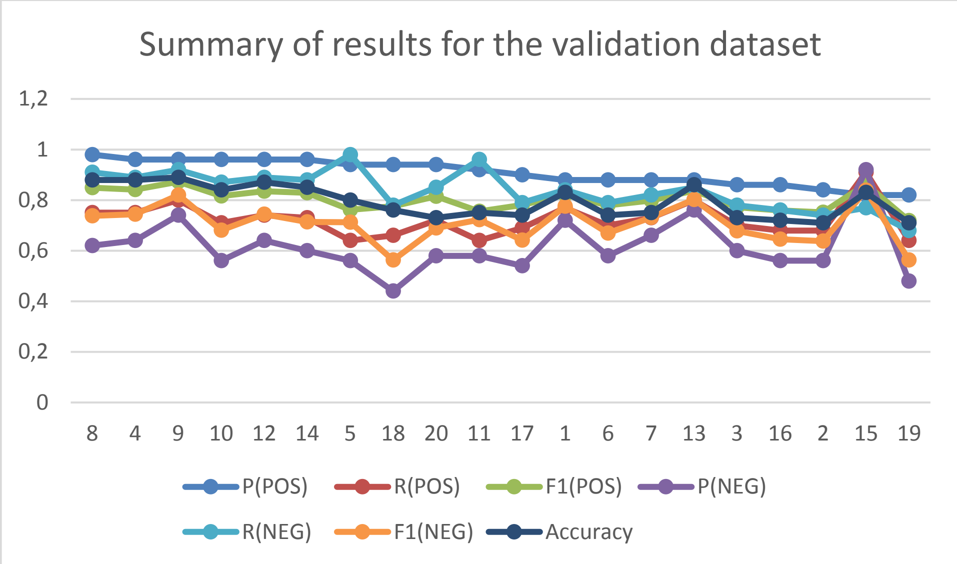

Fig. 4: Summary of experimental results for the LSTM model validation dataset

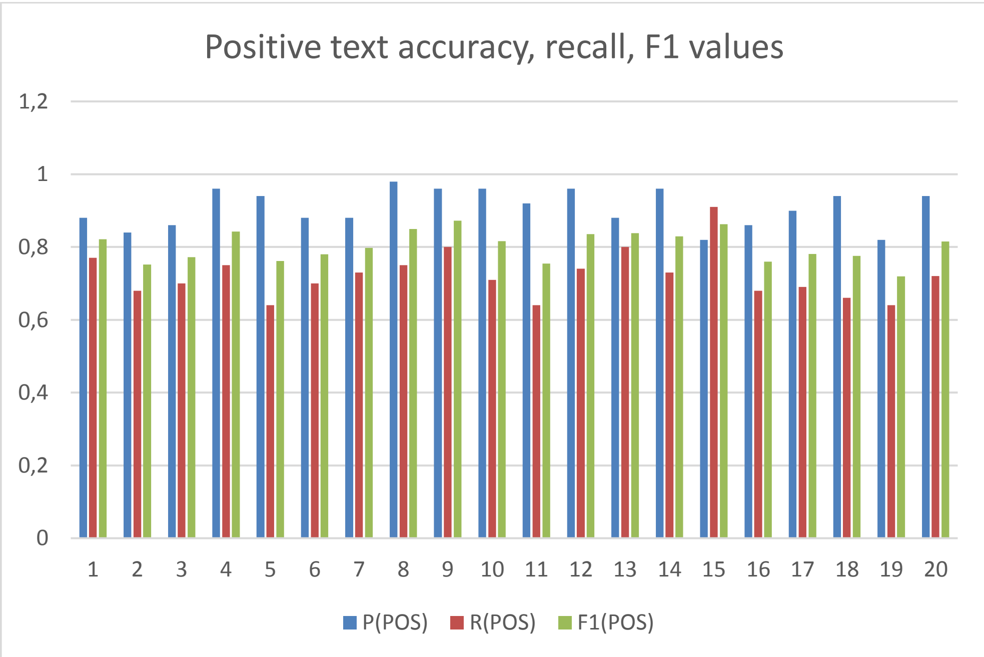

Fig. 5: Accuracy, recall, F1 values for forward text

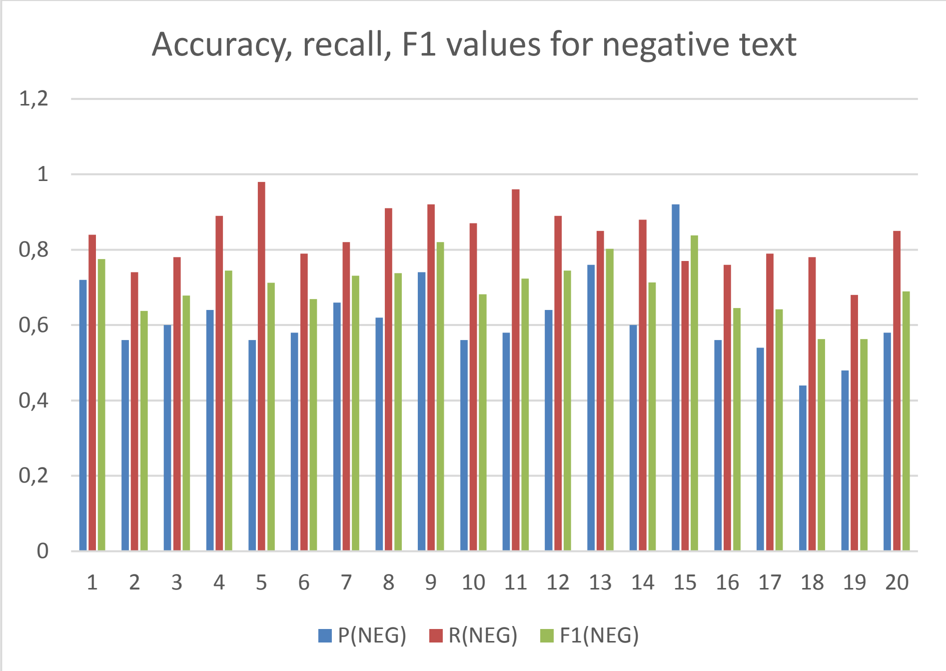

Fig. 6: Accuracy, recall, F1 values for negative text

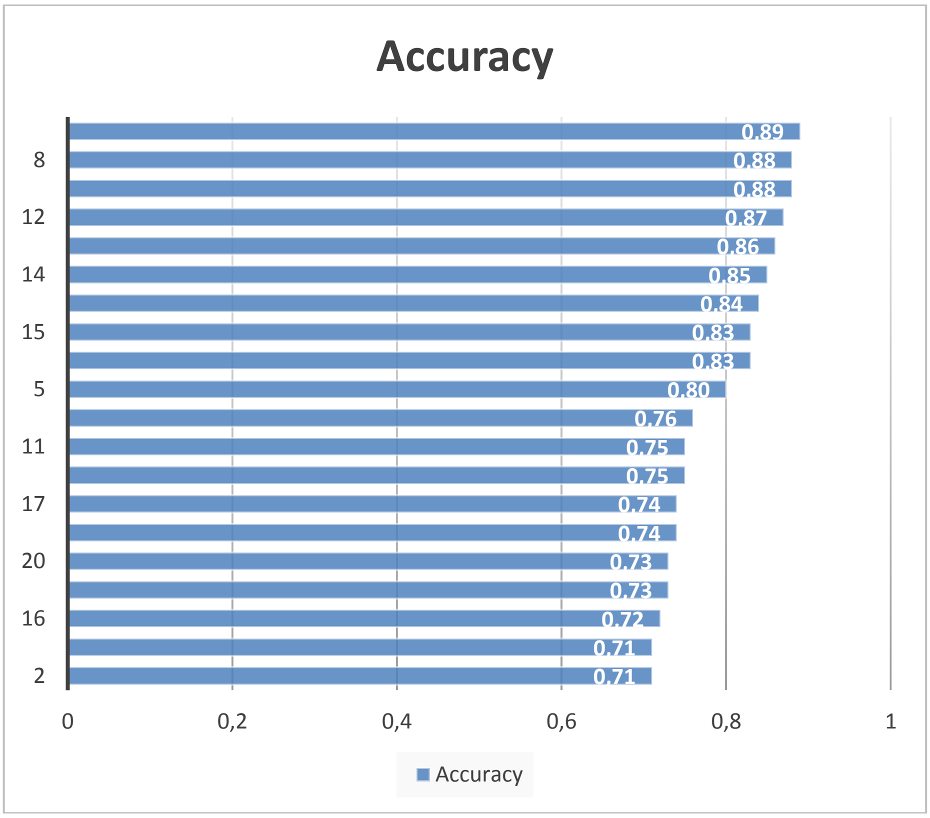

Fig. 7: Accuracy of data sets

Acknowledgment:
We thank the anonymous reviewers and editors for their very constructive comments. This work was supported in part by the Undergraduate teaching quality and teaching reform project of Anhui University of Finance and Economics under Grant No. acszjyyb2021035.

References:
[1] Chaturvedi I, Ragusa E, Gastaldo P, et al. Bayesian network based extreme learning machine for subjectivity detection, Journal of the Franklin Institute, 2017, 355(4): 1780-1797.
[2] Zhou J, Huang J X, Chen Q, et al. Deep learning for aspect-level sentiment classification: survey, vision, and challenges, IEEE Access, 2019,7: 78454-78483.
[3] Thien K T, Thi P T. A hybrid approach for building a Vietnamese sentiment dictionary. Journal of Intelligent & Fuzzy Systems,2018,35(1):1-12.
[4] Wu L, Morestatter F, Liu H, Et al. SlangSD: building, expanding and using a sentiment dictionary of slang words for short text sentiment classification . Language Resources and Evaluation,2018,52:839-852.
[5] Zhang S X, Wei Z L, Wang Y, et al. Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary . Future Generation Computer Systems,2018,81:395-403.
[6] Bravo-Marques F, Khanchandani A, Pfahringer B. Incremental word vectors for time-evolving sentiment lexicon induction. Cognitive Computation,2021,14:425-441.
[7] Asghar M Z, Khan A, Ahmad S, et al. Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS One,2017,12(2): e0171649.
[8] Baid P, Gupta A, Chaplot N. Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications,2017,179(7):45-49.
[9] Hasan A, Moin S, Karim A, et al. Machine learning-based sentiment analysis for twitter accounts. Mathematical and Computational Applications,2018,23(1):11.
[10] Ahmad M, Aftab S, Bashir S, et al. SVM optimization for sentiment analysis. International Journal Advanced Computer Science and Applications,2018,9(4):393-398.
[11] Mathapati S, Nafeesa A, Manjula S H, et al. OTAWE Optimized topic-adaptive word expansion for cross domain sentiment classification on tweets. Advances in Machine Learning and Data Science. Singapore: Springer,2018,705:213-224.
[12] Birjali M, Benihssane A, Erritali M. Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. Procedia Computer Science,2017,113:65-72.
[13] Dwivedi R K, Aggarwal M, Keshari S K, et al. Sentiment analysis and feature extraction using rule-based model (RBM). Proceedings of the 2019 International Conference on Innovative Computing and Communications. Cham: Springer,2019:57-63.
[14] Can E F, Ezencan A, Can F. Multilingual sentiment analysis: an RNN-based framework for limited data [EB/OL]. [2018-06-08]. https://arxiv. org/pdf/1806. 04511. pdf.
[15] Wang Y Q, Huang M L, Zhu X Y, et al. Attention-based LSTM for aspect-level sentiment classification. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics,2016:606-615.
[16] Gopalakrishnan K, Salem F M. Sentiment analysis using simplified long short-term memory recurrent neural networks. https://arxiv. org/abs/2005. 03993v1.
[17] Chen Y X, Yuan J B, You Q Z, et al. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM . Proceedings of the 26th ACM international conference on Multimedia. New York:ACM,2018:117-125.
[18] Chaudhary S, Kakkar M. Sarcasm detection technique on twitter data with natural language processing.Proceedings of International Conference on Big Data, Machine Learning and their Applications. Singapore: Springer,2021:283-293.
[19] Du Y, Li T, Pathan M S, et al. An effective sarcasm detection approach based on sentimental context and individual expression habits. Cognitive Computation,2021(1):1-13.
[20] Zuo E, Zhao H, Chen B, et al. Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. IEEE Access,2020,8:37967-37975.
[21] Chen Peng, Sun Zhongqian. Recurrent attention network on memory for aspect sentiment analysis.Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,2017.
[22] Wei X, Tao L. Aspect based sentiment analysis with gated convolutional networks. arXiv:1805.07043,2018.
[23] Heikal M, Torki M, Elmakky N. Sentiment analysis of arabic tweets using deep learning. Procedia Computer Science,2018,142:114-122.
[24] Al-Smadi M, Talafha B, Al-Ayyoub M, et al. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. International Journal of Machine Learning and Cybernetics,2018,10:2163-2175.

[25] Cai Y, Yang K, Huang D P, et al. A hybrid model for opinion mining based on domain sentiment dictionary. International Journal of Machine Learning and Cybernetics,2019,10:2131-2142.

[26] Graves A, Mohamed A, Hinton G. Speech Recognition with Deep Recurrent Neural Networks.Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2013: 6645-6649.

[27] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM Based Auto-Encoder Model for ECG Arrhythmias Classification", IEEE Transactions on Instrumentation and Measurement, 2020, pp. 1232-1240.

[28] Gregor K, Danihelka I, Graves A, Rezende D J, Wierstral D. DRAW: A Recurrent Neural Network for Image Generation.Proceedings of International Conference on Machine Learning, 2015: 1462-1471.

[29] Mikolov T, Kombrink S, Deoras A, Burget L, Cernocky A J H. RNNLM-Recurrent Neural Network Language Modeling Toolkit.Proceedings of Automatic Speech Recognition and Understanding Workshop, 2011: 196-201.

[30] Minaee S, Kalchbrenner N, Cambria E, et al. Deep learning based text classification: a comprehensive review. Computation and Language, 2020, 8(5): 85616- 85638.

[31] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks. Proceedings of the 31st International Conference on Machine Learning. Beijing: JMLR, 2014: 1764 - 1772.

[32] Akase S, Suzuki J, Nagata M. Input-to-output gate to improve RNN language models. arXiv Preprint, 2017, 2017: arXiv: 1709. 08907.

[33] Miyamoto Y, Cho K. Gated word-character recurrent language model. arXiv Preprint, 2016, 2016: arXiv: 1606. 01700.

[34] Jozefowicz R, Vinyals O, Schuster M, et al. Exploring the limits of language modeling. arXiv Preprint, 2016, 2016: arXiv: 1602. 02410.

[35] CHO K, Van Merrienboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv Preprint, 2014, 2014: arXiv: 1406. 1078.

[36] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv Preprint, 2014, 2014: arXiv: 1409. 0473.

[37] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv Preprint, 2016, 2016: arXiv: 1609. 08144.

[38] Mou, Lichao, Pedram Ghamisi, and Xiao Xiang Zhu. "Deep recurrent neural networks for hyperspectral image classification." IEEE Transactions on Geoscience and Remote Sensing 55.7 (2017): 3639-3655.

[39] Feng Li, Lingling Wang, "Application of Deep Learning Based on Garbage Image Classification," WSEAS Transactions on Computers, vol. 21, pp. 277-282, 2022

[40] Feng Li, Lingling Wang, "Case-Based Teaching for Stock Prediction System Based on Deep Learning," WSEAS Transactions on Business and Economics, vol. 19, pp. 1325-1331, 2022

[41] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 1994, 5(2): 157 - 166.

[42] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735 - 1780.

[43] Gers F A, Schraudolph N N. Learning precise timing with LSTM recurrent networks . Journal of Machine Learning Research, 2002,3(1): 115 - 143.

[44] Gers F A, Schmidhube R J, Cummins F. Learning to forget: continual prediction with LSTM . Neural Computation, 2000, 12(10): 2451 - 2471.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

-Feng Li, Chenxi Cui carried out the simulation and the optimization.

-Yashi Hu, Lingling Wang have organized and executed the experiments of Section 4.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work was supported in part by the Undergraduate teaching quality and teaching reform project of Anhui University of Finance and Economics under Grant No. acszjyyb2021035.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.