future internet

MDPI

Article

BERT- and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords

Xinlu Li, Yuanyuan Lei and Shengwei Ji *

Institute of Applied Optimization, School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China

* Correspondence: swji@mail.hfut.edu.cn

Abstract: Sentiment analysis of online Chinese buzzwords (OCBs) is important for healthy development of platforms, such as games and social networking, which can avoid transmission of negative emotions through prediction of users' sentiment tendencies. Buzzwords have the characteristics of varying text length, irregular wording, ignoring syntactic and grammatical requirements, no complete semantic structure, and no obvious sentiment features. This results in interference and challenges to the sentiment analysis of such texts. Sentiment analysis also requires capturing effective sentiment features from deeper contextual information. To solve the above problems, we propose a deep learning model combining BERT and BiLSTM. The goal is to generate dynamic representations of OCB vectors in downstream tasks by fine-tuning the BERT model and to capture the rich information of the text at the embedding layer to solve the problem of static representations of word vectors. The generated word vectors are then transferred to the BiLSTM model for feature extraction to obtain the local and global semantic features of the text while highlighting the text sentiment polarity for sentiment classification. The experimental results show that the model works well in terms of the comprehensive evaluation index F1. Our model also has important significance and research value for sentiment analysis of irregular texts, such as OCBs.

Keywords: online Chinese buzzwords; sentiment analysis; deep learning; pre-trained language models; BiLSTM

Citation: Li, X.; Lei, Y.; Ji, S. BERT- and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords. *Future Internet* **2022**, *14*, 332. https://doi.org/10.3390/fi14110332

Academic Editor: Tinghuai Ma

Received: 30 September 2022
Accepted: 10 November 2022
Published: 14 November 2022

1. Introduction

Text sentiment analysis, which can extract the sentiment content of texts, is a common application of natural language processing (NLP) and opinion mining. Moreover, sentiment analysis of online Chinese buzzwords (OCBs) is important in the real world. Specifically, sentiment analysis of OCBs (1) helps microblogs, games, and other online platforms to shield against some improper language, (2) purifies the online environment, and (3) precisely grasps the emotional direction of users. OCBs reflect the real life of society spreading through the Internet. OCBs are also are a kind of living language form [1], which is widely used, produced, and applied to the network. OCBs are emerging and changing, and about 1000 OCBs are generated every year and used by millions of users. OCBs are synonymous with Internet language, which refers to a language generated from the Internet or applied to Internet communication, mainly from online games, chat and comments, and other online social platforms. They usually consist of a mixture of Chinese characters, numbers, English letters, and symbols, such as the harmonic word " 鸭梨山大", the superposition " 绝绝子", and the abbreviation " 高大上", etc., often expressing special meanings in specific online media communication. OCBs vary in length, ignore grammar, and have no complete semantic structure [2]. Compared with ordinary text, OCBs can be spread rapidly because they are grounded, personal, creative, short, and interesting [3].

However, OCBs also generate negative effects, such as (1) minors are easily influenced by alienating language forms and develop bad expression habits [4]; (2) OCBs come

in various forms, often have an emotional tendency, and even tend to present vulgarity; and (3) the process of OCBs dissemination often leads to formation of stereotypes in language patterns. It is difficult to form a good social mentality because of the homogenization of people's basic perspectives on social issues or social events and the convergence of their views.

For traditional sentiment analysis models, it is difficult to obtain accurate sentiment characteristics of OCBs. Due their irregular structure and semantics, OCBs have no obvious emotional tendency. Besides, the current main text sentiment analysis models are all based on short text [5]. Furthermore, when the amount of existing datasets is large, the convergence time of the model is often too long.

To solve the above problems, in this paper, the BERT pre-training model and BiLSTM are introduced to learn deep sentiment features from the context of the irregularity and incomplete semantic structure of OCBs. Furthermore, the BERT pre-training model is fine-tuned to accelerate the convergence rate of the model. Motivation to above problems can be seen in Figure 1. The main contributions are summarized as follows.

- The BERT–BiLSTM model is proposed. The model does not require word separation during sentiment analysis and is able to capture deep contextual information from the word order structure. The superiority of BERT–BiLSTM is illustrated in Section 3.
- A fine-tuned sentiment analysis of OCBs is proposed to accelerate the model convergence speed. First, the fine-tuning process directly employs the parameters obtained from the pre-training as the initial values of the proposed model. Then, the manually labeled dataset is transferred according to the downstream tasks to balance the relationship between the data and the model.
- From the extensive experimental results on the OCBs dataset, we can conclude that BERT–BiLSTM outperforms seven state-of-the-art models in terms of recall and F1-score. In addition, the ablation experiments demonstrate the superiority of the proposed model combining BERT and BiLSTM. The proposed model has significant implications for sentiment analysis of OCBs.

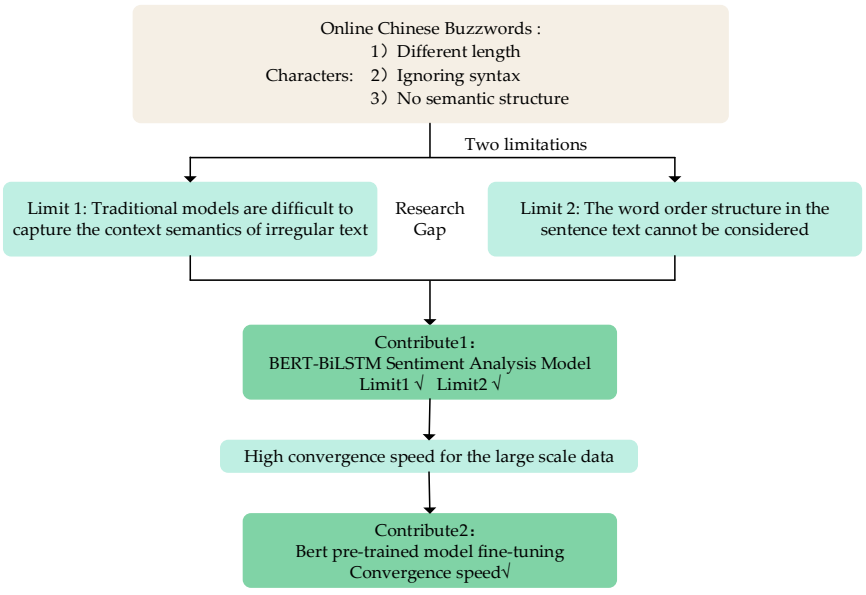

Figure 1. Motivation of sentiment analysis of online Chinese buzzwords.

Although this model has many contributions and advantages, it also has some limitations:

- The proposed model is trained on datasets with small data volume (60,000 online Chinese buzzwords texts) in this paper. We will study its application on large-scale datasets in the future.

- The proposed model is applied to static data in this paper. However, emotional analysis of dynamic data is also of great significance. For example, DTSCM tracks the change trend of theme emotion in different time segments by capturing the theme of microblog messages in different time segments [6].
- This model divides the emotions of online Chinese buzzwords into two categories (i.e., positive and negative), but the classification of multiple categories of emotions needs further research (e.g., sadness, anger, tension, happiness).

The rest of this paper is structured as follows. Section 2 describes the current research in this study. In Section 3, the BERT and BiLSTM models and the proposed hybrid model (BERT–BiLSTM) are described. In Section 4, the performance of the proposed algorithm is analyzed and compared with the performance of several state-of-the-art text classification algorithms. Finally, Section 5 provides a summary and outlook.

2. Related Work

Currently, researchers are focused on studying text sentiment analysis based on sentiment lexicons [7], traditional machine learning [8], and deep learning [9,10]. The sentiment-lexicon-based approach constructs the sentiment lexicon manually and sets the matching rules manually to finally achieve the sentiment analysis of the text. For example, the SentiWordNet [11] sentiment lexicon first combines synonyms based on WordNet and then assigns a positive or negative polarity score to a set of synonyms, which can represent the user's sentiment tendency. Unlike English sentiment lexica, Chinese sentiment lexica are mainly composed of NTUSD [12], the Zhiwang knowledge base, and sentiment vocabulary ontology databases. Wang et al. [13] improved the semantic similarity algorithm based on the knowledge network words, thus improving the semantic similarity accuracy. Hao et al. [14] merged a lexicon of microblog data based on Hownet word similarity and subsequently used pointwise mutual information (PMI) to classify the sentiment of web words. Ye et al. [15] combined the CBOW model and syntactic rules to extract candidate sentiment words from the corpus. They then used the improved SOPMI algorithm to determine the sentiment polarity to form a domain positive- and negative-sentiment lexicon. Collecting and summarizing words with sentiment tendency requires a great deal of time and reading many related materials as well as marking the sentiment polarity and intensity of these words in different degrees. Therefore, the cost of building sentiment dictionaries is large, and sentiment-dictionary-based methods do not consider the relationships between words in the text and lack word sense information.

Machine learning methods show advantages in sentiment analysis tasks compared to sentiment polarity dictionaries. Yang [16] used the TF-IDF algorithm to convert text data into vector data for describing the sentiment of different movie reviews. Three machine learning models were trained by these feature vectors, L1 logistic regression, L2 logistic regression, and a CatBoost model. The results showed that the accuracy and precision of the L1 logistic regression was close to 80%. Tiwari [17] used plain Bayesian (NB) and maximum entropy (ME) methods with a support vector machine (SVM) to perform a review sentiment analysis of tendencies. Since feature selection affects the performance of machine learning methods, Tripathy [18] analyzed online review comments by combining N-gram models with machine learning methods. Their experiments showed that SVM combined with unigram, bigram, and trigram to extract features achieved the best classification results.

With the research and development of deep learning in natural language processing, deep neural networks have achieved outstanding performance in sentiment analysis. Kim [19] solved the sentiment classification problem using convolutional neural networks (CNNs). Cho [20] proposed gated recurrent units (GRUs) to analyze contexts with long dependency problems, which showed significant improvements in various tasks of natural language processing. Qu and Wang [21] proposed a model for sentiment analysis based on hierarchical attention networks with a 5% improvement in accuracy compared to recurrent neural networks. To address the inability of traditional neural networks to fully

capture the entire context of a sentence or comment, some studies have used variants of recurrent neural networks (RNN) [22], such as LSTM or GRU, to solve sentiment analysis problems [23]. LSTM models enable phrase-level sentiment classification to include linguistic regularizations, such as negativity, intensity, and polarity [24]. Nio [25] used bidirectional LSTM to train the labeled text in order to process the syntax and semantics of Japanese. For polarity classification of phrase-level words, Zhang [26] proposed a BiGRU model with multiple inputs and multiple outputs. Thakur [27] studied Twitter sentiments about Omicron variants and analyzed various emotional features, such as "bad", "good", "terrible", and "great." Most tweets are published in multiple languages, which provides some ideas for our next research. Basiri [28] proposed a bidirectional CNN–RNN emotion analysis model based on attention. By taking into account the time information flow to capture the corresponding emotional features, they carried out experiments based on five reviews and three Twitter datasets, with good results.

Considering that a great deal of sentiment analysis is carried out on social media communication, text often ignores grammar and spelling rules, so preprocessing technology is needed to clean up data. Palomino [29] evaluated the effectiveness of different combinations of preprocessing components to obtain the overall accuracy of some existing tools and algorithms.

In the above studies, the sample data basically belong to regular and formal written text data. However, most of the OCBs text data studied in this paper are presented in an irregular form. Since their text does not have an obvious semantic structure, the above CNN and RNN models cannot handle such text well. It is difficult to solve complex emotions in the context of online buzzwords. At the same time, most of the above methods use Word2Vector or GloVe and other static word vector methods. Yet, there are many multiple meanings in OCBs that require richer dynamic word vector expression capability. To address these problems, this paper proposes a BERT pre-training model [30] combined with BiLSTM bi-directional long- and short-term memory networks for OCB sentiment analysis. In this paper, we first generate dynamic word vectors based on the pre-trained BERT model, then transfer the dynamic word vectors into the BiLSTM networks to extract the sentiment features of the text by combining the contextual semantics contained in the word vectors, and finally use the Softmax layer to obtain the sentiment polarity of the current text.

3. Proposed Model

In this section, the proposed BERT–BiLSTM model for sentiment analysis of OCBs is detailed.

3.1. BERT Pretraining Language Model

The BERT model has shown strong advantages in tasks such as text classification and sentiment analysis [30]. Therefore, the BERT model is most likely very suitable for OCBs with strong semantic pertinence and polysemy. BERT is a transfer learning pre-trained neural network model [30]. The Transformer-based bidirectional encoder of this model can take into account the information before and after the word when processing a word so as to obtain the semantics of the context of the word. The basic model of BERT has 12 stacked encoding layers, each encoding layer has 12 self-attention heads, and each feed-forward layer has 768 hidden units, and the output of the final model, which is the input of the downstream task, provides high-quality word vectors [30]. As shown in Figure 2, BERT is pre-trained through a large-scale corpus, which leads to model network parameters suitable for general NLP tasks. It then is fine-tuned to adapt the model to specific downstream tasks.

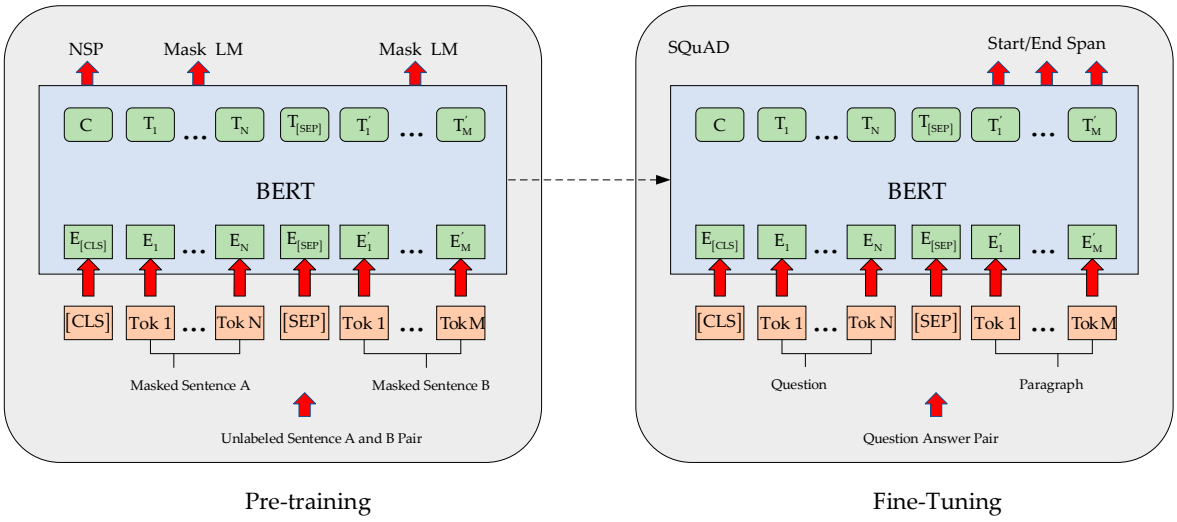

Figure 2. BERT pre-training/fine tuning flow chart.

Since traditional models need input vectors when dealing with natural language, this usually means that the vocabulary and parts of speech need to be converted into sequence features. BERT has advantages over the traditional models, Word2Vec or GloVe [31]. In contrast, BERT can learn word representations based on contextual information and adjusts them according to the meaning of words when fusing contextual information, but words represented by Word2Vec cannot contain context.

The input of the BERT model is represented by the vector superposition of token embedding, segment embedding, and position embedding, and the word vector representation is generated as shown in Figure 3. Token embedding is used as the first word vector marked, and its initial value can be randomly generated, which is the segmentation mark between sentences. Segment embedding is a vector that distinguishes whether different buzzword texts are semantically similar. Position embedding represents location information of each word in the buzzword text. The input vector of the BERT model not only contains short text semantics but also contains the distinction information between different sentences and the position information between words. BERT is pre-trained by using the Masked Language Model (MLM), which randomly hides the input words and then predicts the original vocabulary size of the hidden part according to the context. Next Sentence Prediction (NSP) allows one to insert and label the beginning and end of each sentence, respectively, and predict whether the positions of two sentences are adjacent by learning the relationship between sentences. This bidirectional training model enables BERT to have deeper language context awareness and the ability to learn the context of words through the surrounding environment.

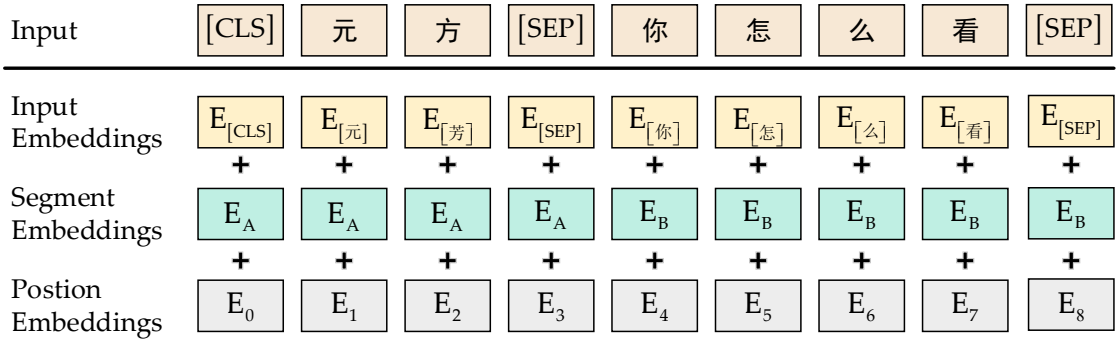

Figure 3. Input characterization diagram of BERT. (The intput " 元芳，你怎么看" is translated as "Yuan Fang, what do you think").

The BERT fine-tuning process directly uses the parameters obtained from pre-training as the initial value of the model, transfers the manually labeled dataset according to the downstream tasks, balances the relationship between the data and the model (hence, the BERT model can be further fitted and converged), and, finally, a model that can be used downstream is obtained.

3.2. BiLSTM Network

A BiLSTM model is a model composed of a forward LSTM and a backward LSTM, and the LSTM in this model is a variant of the recurrent neural network (RNN). In order to solve the problem of gradient disappearance of traditional RNN, a gating unit is introduced into the LSTM, which provides the LSTM with a stronger ability to capture long-term dependencies and enables the RNN to better discover and utilize the dependencies existing in long-distance data.

Each cell unit in LSTM adopts a new structure, which mainly consists of four parts: input gate i_t, output gate o_t, forget gate f_t, and storage unit c_t. The internal structure of a single cell of the LSTM module is shown in Figure 4.

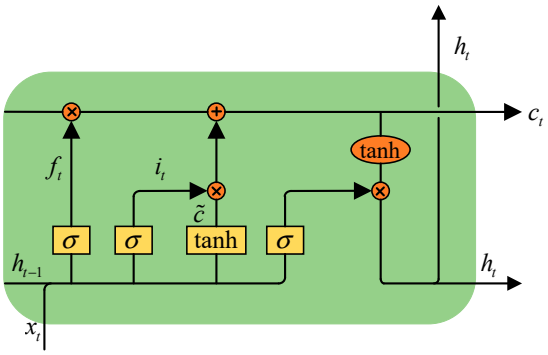

Figure 4. LSTM cell unit.

The LSTM update formula is as follows:

1. Forget gate mechanism:

$$f_t = \sigma(W_{(f)}x_t + U_{(f)}h_{t-1} + b_{(f)}) \tag{1}$$

The forget gate f_t is a reset memory cell. σ represents the Sigmoid activation function, $W_{(f)}$ is a weight matrix, x_t represents the input at time t. U is the weight matrix of the hidden layer output, h_{t-1} is the hidden state, represents the short-term memory. $b_{(f)}$ is the offset vector.

2. Input gate mechanism:

$$i_t = \sigma(W_{(i)}x_t + U_{(i)}h_{t-1} + b_{(i)}) \tag{2}$$

The input gate i_t represents the input gates that control the input of the memory cell. σ represents the Sigmoid activation function, $W_{(i)}$ is a weight matrix, x_t represents the input at time t. U is the weight matrix of the hidden layer output, h_{t-1} is the hidden state, represents the short-term memory. $b_{(i)}$ is the offset vector.

3. Current unit status:

$$\widetilde{c} = f_t \odot c_{t-1} \tag{3}$$

$\widetilde{c}$ is candidate memory cell. The forget gate f_t is a reset memory cell. c_{t-1} is the cell state of $t-1$, represents the long-term memory.

4. Update unit status:

$$c_t = \widetilde{c} + i_t \odot \tanh(W_{(c)} x_t + U_{(c)} h_{t-1} + b_{(c)}) \tag{4}$$

c_t is the cell state, represents the long-term memory, $\widetilde{c}$ is candidate memory cell. i_t represents the input gates that control the input of the memory cell. $W_{(c)}$ is a weight matrix, x_t represents the input at time t. U is the weight matrix of the hidden layer output, h_{t-1} is the hidden state, represents the short-term memory. $b_{(c)}$ is the offset vector.

5. Output gate mechanism:

$$o_t = \sigma(W_{(o)} x_t + U_{(o)} h_{t-1} + b_{(o)}) \tag{5}$$

The output gate o_t represents the output gates that control the output of the memory cell. σ represents the Sigmoid activation function, $W_{(o)}$ is a weight matrix, x_t represents the input at time t. U is the weight matrix of the hidden layer output, h_{t-1} is the hidden state, represents the short-term memory. $b_{(o)}$ is the offset vector.

6. The current state of the hidden layer:

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

h_t is the hidden state. tanh is the Sigmoid activate function. c_t is the cell state, represents the long-term memory. o_t represents the output gates that control the output of the memory cell.

To acquire backward and forward features in a given time, Xu [32] proposed the BiLSTM network. Therefore, in order to solve the problem of feature irregularity in the OCBs, this paper uses the BiLSTM model to capture more features hidden in the contextual deep semantic dependencies. The structure of BiLSTM is shown in Figure 5. Here, BiLSTM is used to learn the output of the BERT layer to enhance the fitting effect of network features and generalization to new datasets.

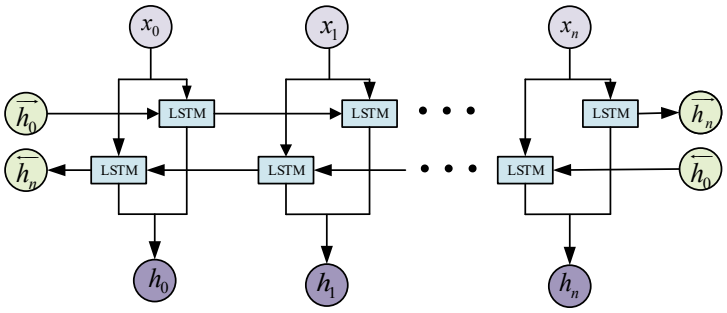

Figure 5. BiLSTM network model.

The formulas for the state at each moment of the model are shown in Equations (7) and (8). The final state is jointly determined by the state of the BiLSTM, as shown in Formula (9) later.

$$\overrightarrow{h_i} = LSTM(x_i, \overrightarrow{h}_{i-1}) \tag{7}$$

$$\overleftarrow{h_i} = LSTM(x_i, \overleftarrow{h}_{i-1}) \tag{8}$$

where x_i represents the input vector at time i, $\overrightarrow{h}_{i-1}$ represents the forward hidden layer vector at time $i-1$, and $\overleftarrow{h}_{i-1}$ represents the reverse hidden layer vector at time $i-1$.

3.3. Network Buzzwords Sentiment Analysis Model

This paper adopts a combined prediction model, namely Bidirectional Encoder Representation of Bidirectional Long Short-Term Memory Transition (BERT–BiLSTM), to con-

struct a sentiment analysis model for OCBs. It is worth noting that the BERT–BiLSTM proposed in this paper uses BERT as the upstream module and BiLSTM as the downstream module to fine-tune the BERT pre-trained model and then transfer it into BiLSTM for training. The BERT–BiLSTM OCBs structure is shown in Figure 6.

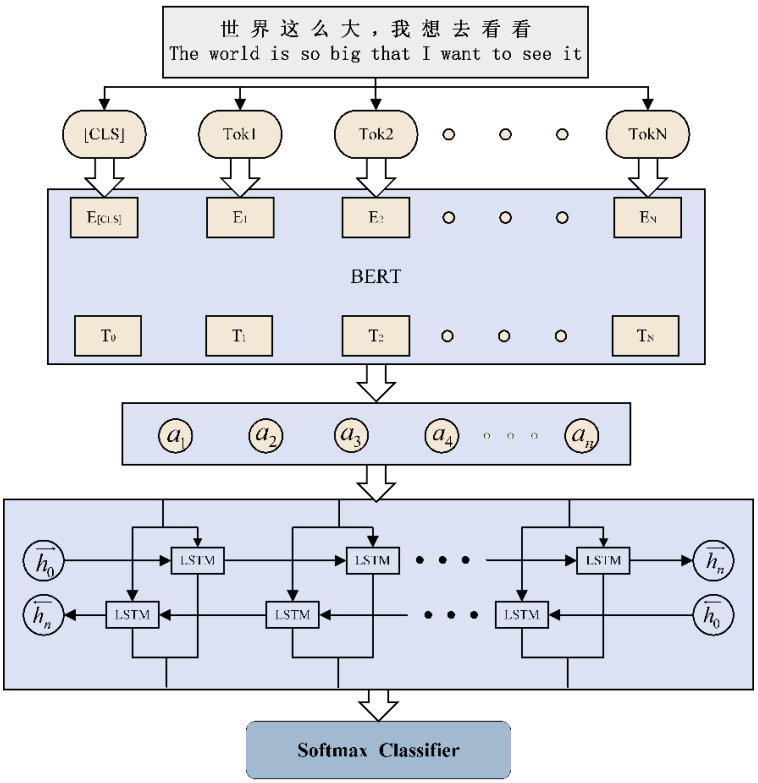

Figure 6. Structure of BERT–BiLSTM network buzzwords.

As mentioned above, OCBs have some problems, such as language fragmentation, incomplete semantic structure, and irregular features, while BERT has a strong ability to learn the features of nearby words. In Figure 6, $E_1, E_2, \cdots, E_N$ is the word input vector of the BERT layer, $E_{[CLS]}$ is the buzzword start bit indicator, T_0 is the output vector of the OCBs start symbol after BERT training, and $T_1, T_2, \cdots, T_N$ is the word output vector of the BERT layer. Subsequently, a BiLSTM model is used to extract contextual features in the input sequence data, where $\overrightarrow{h}_1, \overrightarrow{h}_2, \cdots, \overrightarrow{h}_n$ is the state of the forward LSTM hidden layer of the BiLSTM layer, $\overleftarrow{h}_1, \overleftarrow{h}_2, \cdots, \overleftarrow{h}_n$ is the state of the reverse LSTM hidden layer of the BiLSTM layer, respectively. Finally, the feature vector output by BiLSTM is analyzed by a Softmax classifier to achieve the sentiment polarity classification of OCBs.

The proposed BERT–BiLSTM model multiplies the output $C = \{T_0, T_1, T_2, \cdots, T_N\} \in R^{\mathrm{n}}$ (including $[CLS]$) of the hidden layer training of BERT by the learnable weight $W_a \in R^{d_a \times n}$ as the input of the BiLSTM model. The formula is as follows:

$$a_i = g_1(W_a C_i + b_a) \tag{9}$$

where $1 \leq i \leq n$, n denotes the dimension of the feature vector output after fine-tuning of the BERT model, $a_i \in R^{da}$ is the input vector of the BiLSTM layer, b_a is the offset vector of dimension d_a. Here, we adopt Sigmoid as the activation function g_1.

The ordinary LSTM model calculates the one-way hidden layer sequence h, while BiLSTM calculates the forward hidden layer vector as $\overrightarrow{h}$ and the backward hidden layer

vector as $\overleftarrow{h}$ and finally combines the two for output vector v_i, as shown in Figure 7. The calculation formula is as follows:

$$v_i = \overrightarrow{h_i} + \overleftarrow{h_i} \tag{10}$$

where $\overrightarrow{h_i} \in R^{d_h}$,$\overleftarrow{h_i} \in R^{d_h}$.

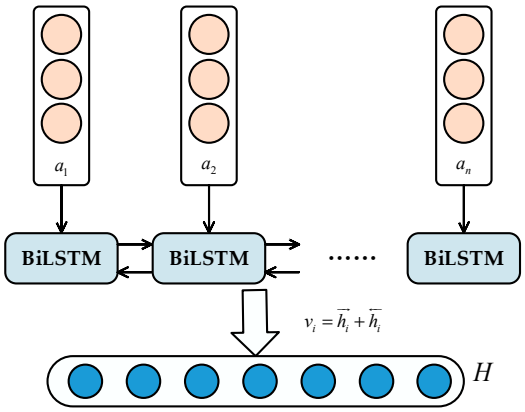

Figure 7. BiLSTM layer language model structure.

In addition, we define the hidden layer of the model as follows:

$$h_i^d = g_2\left(W_h^d a_i + U h_{i-1}^d + b_h^d\right) \tag{11}$$

where g_2 is the Tanh function, $W_h^d a_i \in R^{d_h \times d_a}$ is the weight matrix of a_i, $d \in \{0, 1\}$ represents two different directions in the hidden layer, U is the weight matrix of the hidden layer output sequence h^d at moment i. h_{i-1}^d corresponds hidden layer output sequence of the previous moment $i-1$; b_h represents offset vector of d direction. Then, the output sequences h_d of all hidden layers are combined to obtain the feature vector H at the end of the sentence. Then, the feature vector H is transferred to the fully connected layer with the Relu function and the Softmax function to classify the emotional tendencies. The probability calculation formula of sentiment tendency classification is as follows:

$$p(y|H, W_s, b_s) = \text{softmax}(W_s H + b_s) \tag{12}$$

where $W_s \in R^{|s| \times |l|}$ ($|l|$ is the number of classes) and $b_s \in R^{|l|}$ are the learnable parameters of the output layer.

4. Experiments and Analysis

4.1. Datasets

The dataset in the experiments is based on the Internet corpus of Sogou Labs [33] and Weibo popular events [34]. A total of 100,000 OCBs are collected. About 10,000 texts are not appropriate for the experiments, such as being either too long, having complex emotional tendencies, or too many special symbols. Therefore, the original text data are preprocessed and filtered. Finally, there are 90,000 OCBs that meet the criteria of the input model, half of the positive data and half of the negative data. The code and dataset (https://github.com/Rachel-loo/BERT-BiLSTM, accessed on 7 November 2022) have been published to facilitate follow-up research by other researchers. Most of the special texts, such as OCBs, do not carry the labeling of emotionally inclined words. In order to ensure the validity of the data, the text content is first manually labeled and classified (the labeling is divided into two categories: negative and positive). In view of the dichotomy of sentiment research in this paper, negative OCBs are represented by 0 and positive OCBs are represented by 1. We selected 80% of the dataset as the training set, 10% as the test set, and 10% as the validation set (see Table 1).

Table 1. Collection of online Chinese buzzwords.

Text of Online Chinese Buzzwords	Label
坐沙发，赏灯会。 Sit on the couch, enjoy the lights.	1
这是什么神仙操作。 What kind of celestial manipulation is this.	1
悲催！隐形眼镜配戴时间过长，眼睛又发炎了。 Grief! Time of contacting lenses is too long. The eyes are inflamed. again.	0
亲爱滴，有点像手表广告，不过脸很美。 Dear drip, it's a bit like a watch advertisement, but the face is beautiful.	1
我的hotmail 邮箱被黑了，莫名其妙自动给所有人发了个链接。 My hotmail email was hacked and somehow automatically sent a link. to everyone	0
头晕晕，眼花花，早上醒不来，各种状态不佳，我不能再这样下去了。 Dizziness, dizziness, not waking up in the morning, all kinds of poor state, I can't go on like this.	0

4.2. Parameters

All experiments are executed under Ubuntu 20.04 with a GeForce RTX 3080 (Nvidia, Santa Clara, CA, USA). The experimental framework is built by PyTorch (Torch1.11.0, Python 3.7).

The dynamic learning rate and early stopping in the BERT model can determine the parameters, e.g., num_epochs and learning_rate. This study sets num_epochs = 9 to indicate that the model has been trained 9 times because, after testing multiple epochs, the results of each metric set to 9 are the best. If the current accuracy is not improved compared to the previous epoch, reduce the current learning rate. If no improvement happens during 9 epochs, the training is stopped, and the learning rate of the model is 5×10^{-5}. Batch_size is the number of training samples in each batch. When Batch_size = 16, 32, 64, as the number of iterations increases, the training speed becomes slower. The larger the Batch_size, the better the characteristics of the entire data can be displayed and the more accurate the gradient descent direction can be ensured. When set to 256, the number of iterations is reduced, but the parameter correction is slow. Finally, the value of Batch_size is set to 128. Pad_size indicates the length of each text processing; short text is filled and long text is divided. Set the value of Pad_size to 64 because the Internet catchphrase text is short text. In the BERT model [35], Hidden_size is set to 768, indicating the number of neurons in the hidden layer of the model, which is not changed in the combined model, and Hidden_size is still set to 768 (see Table 2).

Table 2. Experimental parameter setting.

Parameters	Values
Num_epochs	9
Batch_size	128
Pad_size	64
Hidden_size	768
Learning_rate	5×10^{-5}
Filter_sizes	(2, 3, 4)

4.3. Metrics

In this paper, precision, recall, and the F1-score are used as evaluation metrics. The precision represents the proportion of positive samples that are correctly predicted. The recall rate stands for the ability of recognizing the positive samples. If the positive and negative datasets are irregular text, there will be gaps in the calculation of precision and recall. The F1-score combines with the two metrics of precision and recall to more comprehensively reflect the classification performance. The better the performance of the classifier, the closer the F1-score is to 1. The specific form is detailed as follows:

Precision is defined as:

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

Recall is defined as:

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

The F1-score is defined as:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{15}$$

4.4. Experiment Analysis

To further evaluate the proposed BERT–BiLSTM OCBs sentiment analysis model, we set up eight sets of experiments. The values of precision, recall, and F1 are obtained on the test set, respectively. The comparison results of the constructed OCBs dataset are shown in Table 3.

Table 3. Evaluation indicators.

Result	Positive	Negative
Guessed pos	TP	FP
Guessed neg	FN	TN

According to Table 4, the results of the proposed BERT–BiLSTM model on the OCBs dataset are better than those of the other models. First, the F1-score of the BERT–BiLSTM model is the highest, which means that the model has strong comprehensive ability and good generalization performance. Second, the recall value of the BERT–BiLSTM model is also the highest compared to other models. This also shows that the number of positive samples identified by the model is the best, and the coverage of training samples is wide. It reflects the sensitivity of the model. Third, in terms of precision, BERT–BiLSTM performs well. Its value is close to BERT, slightly (only 0.54%) worse than BERT–CNN and BERT–RNN. This is due to the high recall rate of the model, which indirectly leads to a decrease in the accuracy rate. Notably, the ability of the BiLSTM model to learn text context features is more prominent, so the results of the three models, BERT–BiLSTM, BiLSTM, and BiLSTM–Attention, are better in terms of recall and F1-score values. The BERT–BiLSTM model has the best training results. This proves that the word vector obtained by using the BERT pre-training model contains more complete context information, which is conducive to extraction of text information by the subsequent model.

Table 4. Model comparison.

Models	Precision (%)	Recall (%)	F1-Score (%)
ERNIE [36]	94.20	93.28	93.74
BiLSTM–Attention [37]	93.59	93.63	92.61
BiLSTM [37]	92.05	92.96	92.50
TextCNN [38]	92.83	91.77	92.29
BERT–BiLSTM	93.78	**93.80**	**93.79**
BERT–CNN	**94.32**	92.81	93.56
BERT–RNN	94.27	93.00	93.63
BERT [30]	93.86	93.55	93.71

Figure 8 demonstrates the comparison of F1 score in the eight sets of experiments. The F1-score of the BERT–BiLSTM model is the highest among the eight sets of comparative experiments. It also shows the improvement in the overall classification performance of the model and the improvement in the ability to distinguish different categories, which proves that the model is performing strong emotional analytical skills.

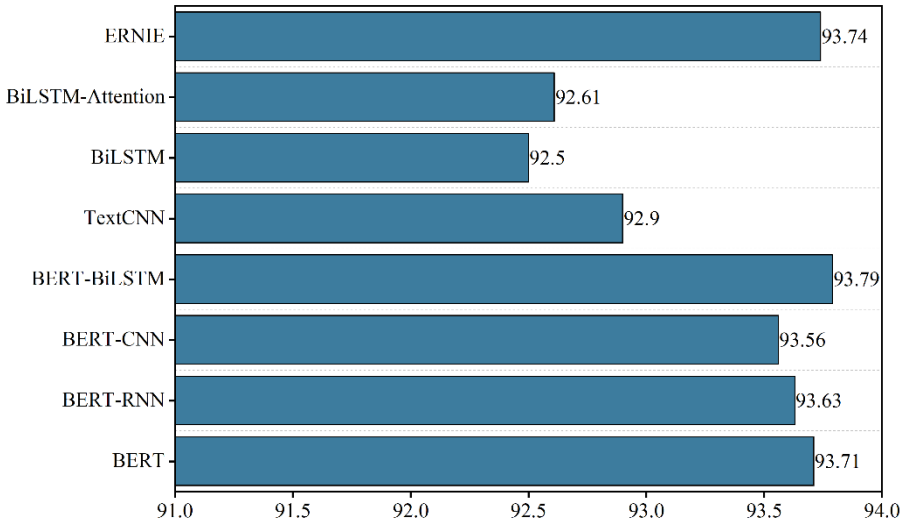

Figure 8. F1-score comparison chart for each experiment.

In order to verify the stability of the model, we compared the performance of the model's accuracy (Acc) and loss in the training set and verification set at different iterations. Comparing Figures 9 and 10 shows that the model has strong stability and good training effect. The final validation (Val) Loss and Val Acc on the test set are 0.15 and 93.74%, respectively.

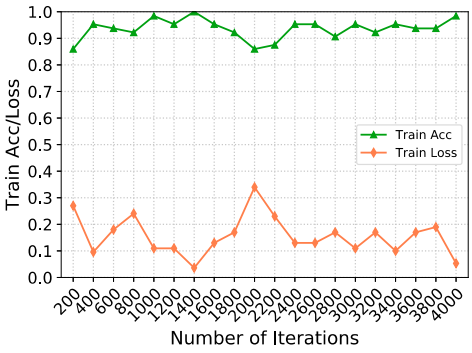

Figure 9. Train Acc/Loss graph.

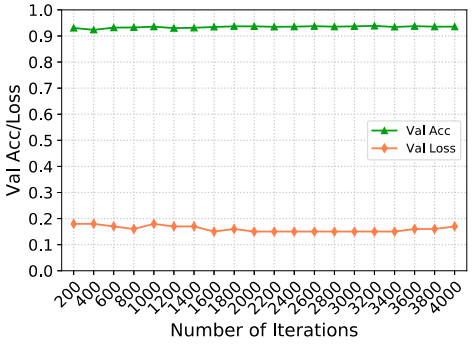

Figure 10. Val Acc/Loss graph.

Ablation Experiment

The proposed model uses BERT pre-training model and BiLSTM model. Because the experimental results show that the BERT–BiLSTM combined model has a better effect on emotion analysis than the above two models, it is compared with these two models in the ablation experiment [39] to prove the superiority. In this study, the following three models for emotional polarity classification were benchmark-tested because they achieved

good results. In order to maintain fairness, the parameters are consistent with the previous experiments. The three models are as follows:

- BERT: The word vector output from BERT pre-training model is used to calculate the emotional polarity through Softmax (a traditional classifier).
- BiLSTM: The OCBs vector is generated through word2vec (a typical method of word embedding), then the generated word vector is input into the BiLSTM model for feature extraction, and then emotional classification is conducted through Softmax (a traditional classifier).
- BERT–BiLSTM: By fine-tuning the BERT model, the dynamic representation of the OCBs vector is generated in the downstream task, then the generated word vector is input to the BiLSTM model for feature extraction, and then the emotional classification is conducted through Softmax (a traditional classifier).

In order to prove the superiority of proposed BERT–BiLSTM model, ablation experiments are conducted on the OCBs dataset (as Table 5). First, the results show that the model has a more obvious boosting effect in terms of recall and F1 values. Second, in terms of precision value, the training effect of single BERT model is slightly better, which indicates that, for such irregular texts as OCBs, it is more necessary to consider the semantic relationship between contexts to obtain more accurate sentiment tendency.

Table 5. Ablation experiment.

Models	Precision (%)	Recall (%)	F1-Score (%)
BERT [30]	**93.86**	93.55	93.71
BERT–BiLSTM	93.78	**93.80**	**93.79**
BiLSTM [37]	92.05	92.96	92.50

BERT model and BERT–BiLSTM model fine-tune downstream tasks by pre-training model. According to the downstream task, the dataset was input manually to balance the relationship between the data and the model so that the model could be further fitted and converged so as to improve the convergence speed of the model. In general, BERT and BERT–BiLSTM converge faster than BiLSTM. As shown in Figure 11, after the fourth iteration, a relatively stable value is obtained, which is due to the advantages of BERT pre-training model. According to the results of F1 and recall values, BERT–BiLSTM model has the best comprehensive ability and superiority.

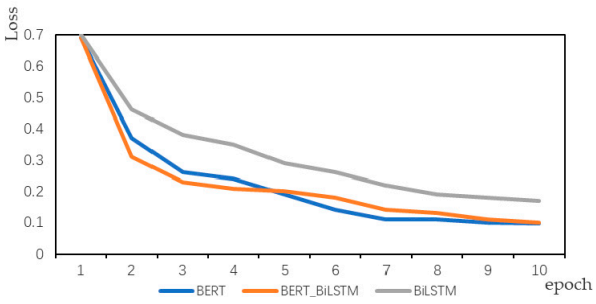

Figure 11. Comparison of convergence speed.

5. Conclusions

The widespread use of social media platforms has brought about an endless stream of new expressions. A large number of OCBs without complete semantic structure and obvious emotional characteristics have emerged. We collected popular OCBs that have been widely used in the past ten years as the research object. We introduced suitable BERT and BiLSTM models and proposed a new method for feature extraction and sentiment orientation classification of OCBs. We solved the problem that the model word vector cannot cover the contextual semantic information. It is difficult to deal with irregular and

informally written texts. We used the BERT model to pre-train on OCBs, obtained the feature representation of the input text, and then used it as a bidirectional LSTM. The input of the model is used for sentiment classification training. The model has achieved excellent results in the experiments on the network catchphrase text dataset.

In the future, we will expand the scope of data for in-depth research. We will analyze a positive and negative emotion index so as to provide better methods and suggestions for improving the popularity of Internet sentiment language analysis. This is because the amount of data collected is not yet large enough to mine more emotional features for analysis, so the degree of negative words cannot yet be refined.

Author Contributions: Conceptualization, X.L., Y.L. and S.J.; Funding acquisition, X.L.; Project administration, X.L.; Supervision, X.L.; Writing—original draft, Y.L.; Writing—review & editing, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NSFC under Grant No. 62176085 and Graduate Innovation Project of Hefei University under Grant No. 21YCXL20.

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zan, H.Y.; Xu, H.F.; Zhang, K.L. The construction of Internet slang dictionary and Its analysis. *J. Chin. Inf. Process.* **2016**, *30*, 133–139.
2. Cheng, Y. A study on the standardization of modern Chinese from the perspective of Network language Niche. *This Anc. Invasive* **2022**, *12*, 126–128.
3. Tang, L. A study on the dissemination influence of contemporary Chinese internet buzzwords—Taking 15 internet buzzwords in the first half of 2015 as an example. *J. Hubei Univ. Natl. (Soc. Sci. Ed.)* **2016**, *34*, 139–144.
4. Ji, W. An analysis of the youth mentality behind internet buzzwords. *People's Trib.* **2022**, *4*, 28–31.
5. Liu, W.; Li, Y.; Luo, J. Sentiment analysis of Chinese short text based on BERT and BiLSTM. *J. Taiyuan Norm. Univ. Nat. Sci. Ed.* **2020**, *19*, 52–58.
6. Li, C.; Huang, F.; Wen, X.; Li, X.; Yuan, C.A. Evolution analysis method of microblog topic-sentiment based on dynamic topic sentiment combining model. *J. Comput. Appl.* **2015**, *35*, 2905–2910.
7. Zhang, S.; Wei, Z.; Wang, Y.; Liao, T. Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Future Gener. Comput. Syst.* **2018**, *81*, 395–403. [CrossRef]
8. Gang, Z.; Zan, X. Research on the sentiment analysis model of product reviews based on machine learning. *Comput. Eng. Appl.* **2017**, *3*, 166–170.
9. Tang, L.; Xiong, C.; Wang, Y.; Zhou, Y.; Zhao, Z. Review of deep learning for short text sentiment tendency analysis. *J. Front. Comput. Sci. Technol.* **2021**, *15*, 794–811.
10. Wang, T.; Yang, W. Review of text sentiment analysis methods. *Comput. Eng. Appl.* **2021**, *57*, 11–24.
11. Madani, Y.; Erritali, M.; Bengourram, J.; Sailhan, F. A Hybrid Multilingual Fuzzy-Based Approach to the Sentiment Analysis Problem Using SentiWordNet. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2020**, *28*, 361–390. [CrossRef]
12. Ku, L.; Chen, H.-H. Mining opinions from the Web: Beyond relevance retrieval. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1838–1850. [CrossRef]
13. Wang, H.; Marius, P.; Pan, J. Research on improved algorithm of word semantic similarity based on HowNet. *Comput. Digit. Eng.* **2022**, *50*, 225–228.
14. Hao, M.; Chen, L. Chinese Microblog polarity classification based on Hownet and PMI. *J. Electron. Sci. Technol.* **2021**, *34*, 50–55.
15. Ye, X.; Cao, J.; Xu, F. Sentiment dictionary adaptive learning method in Chinese domain. *Comput. Eng. Des.* **2020**, *41*, 2231–2237.
16. Yang, Z. Sentiment Analysis of Movie Reviews based on Machine Learning. In Proceedings of the 2th International Workshop on Artificial Intelligence and Education, Montreal, QC, Canada, 6–8 November 2020; pp. 1–4.
17. Tiwari, P.; Mishra, B.K.; Kumar, S.; Kumar, V. Implementation of n-gram Methodology for Rotten Tomatoes Review Dataset Sentiment Analysis. *Int. J. Knowl. Discov. Bioinform.* **2017**, *7*, 689–701. [CrossRef]
18. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **2016**, *57*, 117–126. [CrossRef]
19. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv preprint* **2014**, arXiv:1408.5882.
20. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint* **2014**, arXiv:1409.1259.
21. Qu, Z.; Yuan, W.; Wang, X. A Transfer Learning Based Hierarchical Attention Neural Network for Sentiment Classification. In Proceedings of the International Conference on Data Mining & Big Data, Belgrade, Serbia, 14–20 July 2020; Springer: Cham, Switzerland, 2018; pp. 383–392.

22. Abid, F.; Alam, M.; Yasir, M.; Li, C. Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. *Future Gener. Comput. Syst.* **2019**, *95*, 292–308. [CrossRef]
23. Arkhipenko, K.; Kozlov, I.; Trofimovich, J. Comparison of neural network architectures for sentiment analysis of Russian tweets. *Comput. Linguist. Intellect. Technol. Proc. Int. Conf. Dialogue* **2016**, *15*, 50–59.
24. Qian, Q.; Huang, M.; Lei, J. Linguistically regularized lstms for sentiment classification. *arXiv preprint* **2016**, arXiv:1611.03949.
25. Nio, L.; Murakami, K. Japanese sentiment classification using bidirectional long short-term memory recurrent neural network. In Proceedings of the Japanese Sentiment Classification Using Bidirectional Long Short-Term Memory Recurrent Neural Network, Okayama, Japan, 13–15 March 2018; pp. 1119–1122.
26. Zhang, M.; Zhou, Z. A Review on Multi-Label Learning Algorithms. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1819–1837. [CrossRef]
27. Thakur, N.; Han, C.Y. An Exploratory Study of Tweets about the SARS-CoV-2 Omicron Variant: Insights from Sentiment Analysis, Language Interpretation, Source Tracking, Type Classification, and Embedded URL Detection. *COVID* **2022**, *2*, 1026–1049. [CrossRef]
28. Basiri, M.E.; Nemati, S.; Abdar, M.; Cambria, E.; Acharyad, U.R. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Gener. Comput. Syst.* **2021**, *115*, 279–294. [CrossRef]
29. Palomino, M.A.; Aider, F. Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis. *Appl. Sci.* **2022**, *12*, 8765–8786. [CrossRef]
30. Devlin, J.; Chang, M.-W.; Lee, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint* **2018**, arXiv:181004805.
31. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar; 2014; pp. 1532–1543.
32. Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; Wu, X. Sentiment analysis of comment texts based on BiLSTM. *IEEE Access* **2019**, *7*, 51522–51532. [CrossRef]
33. Internet Corpus of Sogou Labs. Available online: https://pinyin.sogou.com/dict/search/search_list/%CD%F8%C2%E7%C1%F7%D0%D0%D3%EF/normal (accessed on 27 September 2022).
34. Weibo Popular Events. Available online: https://weibo.com/a/hot/realtime (accessed on 27 September 2022).
35. Alhaj, Y.A.; Dahou, A.; Al-qaness, M.A.; Abualigah, L.; Abbasi, A.A.; Almaweri, N.A.O.; Elaziz, M.A.; Damaševičius, R. A Novel Text Classification Technique Using Improved Particle Swarm Optimization: A Case Study of Arabic Language. *Future Internet* **2022**, *14*, 194. [CrossRef]
36. Wang, H.; Sun, M. Chinese short text classification based on ERNIE-RCNN model. *Comput. Technol. Dev.* **2022**, *32*, 28–33.
37. Ge, H.; Zheng, S.; Wang, Q. Based BERT-BiLSTM-ATT Model of Commodity Commentary on The Emotional Tendency Analysis. In Proceedings of the 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI), Qingdao, China, 2–4 July 2021; pp. 130–133.
38. Ce, P.; Tie, B. An analysis method for interpretability of CNN text classification model. *Future Internet* **2020**, *12*, 228. [CrossRef]
39. Meyes, R.; Lu, M.; de Puiseau, C.W. Ablation studies in artificial neural networks. *arXiv preprint* **2019**, arXiv:1901.08644.