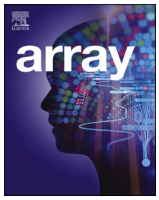
Array 25 (2025) 100378

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array



Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model

Dianwei Chi [a,*], Tiantian Huang [a], Zehao Jia [a], Sining Zhang [b]

[a] School of Artificial Intelligence, Yantai Institute of Technology, Yantai, 264000, China
[b] Department of Oriental Studies, Belarusian State University, Minsk, Belarus

ARTICLE INFO

Keywords:
BERT
TCN
BiLSTM
Attention
Sentiment analysis
NLP
Semantic coding

ABSTRACT

Due to the high semantic flexibility of Chinese text, the difficulty of word separation, and the problem of multiple meanings of one word, a sentiment analysis model based on the combination of BERT dynamic semantic coding with temporal convolutional neural network (TCN), bi-directional long- and short-term memory network (BiLSTM), and Self-Attention mechanism (Self-Attention) is proposed. The model uses BERT pre-training to generate word vectors as model input, uses the causal convolution and dilation convolution structures of TCN to obtain higher-level sequential features, then passes to the BiLSTM layer to fully extract contextual sentiment features, and finally uses the Self-Attention mechanism to distinguish the importance of sentiment features in sentences, thus improving the accuracy of sentiment classification. The proposed model demonstrates superior performance across multiple datasets, achieving accuracy rates of 89.4 % and 91.2 % on the hotel review datasets C1 and C2, with corresponding F1 scores of 0.898 and 0.904. These results, which surpass those of the comparative models, validate the model's effectiveness across different datasets and highlight its robustness and generalizability in sentiment analysis. It also shows that BERT-based coding can improve the model's performance more than Word2Vec.

1. Introduction

With the ripening of mobile Internet technology, more and more data are being generated by users through mobile networks, resulting in a high rate of growth in the volume of users. Due to the existence of mobile terminals, the transmission of information is no longer limited by time and space, and Internet big data experiences geometric growth. Among them, a considerable part of this data comes from social media. Internet users are accustomed to posting their subjective opinions and perspectives on social media, and commentary is itself emotionally attached. Statistical analysis of its totality and access to the users' emotional tendencies and attitudes can offer effective support for relevant management departments and enterprises to understand public opinions and to formulate relevant measures. Hotel reviews are the subjective views and opinions expressed by users on the multi-dimensional accommodation experience, such as hotel service and setting, which, as a part of public resources, can play a certain role in the decision-making of potential customers. More and more online review companies are getting optimization from analyzing customers' online reviews, which can enhance customer loyalty while improving customer experience. Review and recommendation systems can significantly reduce the search cost of potential users, and the processing of unstructured text would help better recognize the factors influencing customer satisfaction, leverage the value of online reviews, and improve the quality of their platforms. The customer reviews on hotels are explicit, short, and topical, and much work must be done to explore better the viewpoints hidden in all this noise.

Sentiment analysis techniques can extract users' subjective emotions and satisfaction from massive text datasets, providing important references for applications such as opinion mining and monitoring [1,2]. Text sentiment analysis mainly adopts three methods: sentiment dictionary, machine learning, and deep learning. Sentiment dictionary methods rely on high-quality dictionaries for sentiment categorization but have high maintenance costs. On the contrary, machine learning methods utilize labeled datasets to train classification algorithms. In contrast, the emergence of deep learning methods compensates for the limitations of these methods and provides advanced features for natural language processing.

Deep learning, especially Transformer-based models, has emerged as a powerful approach to sentiment analysis. These models utilize contextual embedding to enhance the understanding of textual semantics. This feature addresses challenges such as polysemy and word

* Corresponding author.

https://doi.org/10.1016/j.array.2025.100378
Received 1 July 2024; Received in revised form 6 February 2025; Accepted 19 February 2025
Available online 19 February 2025

separation in languages with high semantic flexibility, such as Chinese. Recent advances in Transformer models have revolutionized the field, improving the accuracy and efficiency of sentiment classification tasks. Recent advances in Transformer-based models (especially the BERT model) have revolutionized text mining and sentiment analysis by enabling dynamic word encoding based on contextual semantics.

In this paper, we propose a new model for sentiment analysis that combines TCN, BiLSTM, and Self-Attention mechanism. BERT serves as the contextual embedding backbone, where its deep bidirectional Transformer layers effectively encode subtle semantic nuances and pragmatic polarity shifts in texts, and then sentiment classification accuracy is improved by effectively optimizing the feature vectors through advanced feature extraction using TCN and contextual feature extraction using BiLSTM. TCN is able to efficiently capture long-distance dependencies through parallel processing, while BiLSTM is able to capture bi-directional contextual information, which is particularly useful for sequential text modeling. TCN's dilated convolutions effectively capture long-range dependencies through exponentially expanding receptive fields, they may underutilize local bidirectional sequential patterns due to their unidirectional causal structure. BiLSTM explicitly models forward and backward temporal interactions through dual-gated memory cells, demonstrating superior capability in learning local contextual dynamics. The combination of the two can make the feature extraction more adequate and efficient, which is conducive to improving the prediction effect of the combined model. Self-Attention dynamically weights emotionally salient lexical units, overcoming the positional bias inherent in purely convolutional or recursive methods, for focusing on important words in text.

2. Related research

Among the traditional methods of text sentiment analysis, two methods are used more often. The first method is based on a sentiment dictionary. The sentiment dictionary method can be used to judge the sentiment category by matching the text to be analyzed with the words in the dictionary, and the sentiment value of the sentence would be obtained after calculation; this method requires a relatively high-quality dictionary, and the maintenance and computational expenses are very high. The second is the machine learning approach. Machine learning methods use manually labeled text data and labels combined with machine learning algorithms such as Support Vector Machines, Naïve Bayes, and other machine learning algorithms to achieve text sentiment classification. As in the paper [3], a plain Bayesian-based sentiment classification model is constructed based on a polynomial Bayesian classifier. The study is based on movie review data for training and prediction, and experiments show that the model has good prediction results. Chang [4] made an experimental comparison by applying a support vector machine and plain Bayesian classifier, which illustrated the better performance of SVM in dealing with text sentiment analysis problems. Zhu et al. [5] will conduct experiments based on Tibetan microblogging data collected on Sina Weibo with real-meaning word extraction. Finally, the SVM algorithm is utilized for sentiment analysis, and the training efficiency and accuracy of the model are improved. Chen et al. [6] proposed a PSO-SVM sentiment analysis model using a nonlinear function method to improve the particle swarm algorithm. The model is better than a single model in the experiments on the sentiment classification of movie review data, which alleviates the phenomenon of particles easily falling into local optimality to a certain extent and improves the prediction accuracy. The classification effect of machine learning-based methods depends on the quality and quantity of manual labeling, which is a huge amount of work. Deep learning-based sentiment analysis is proposed to overcome the shortcomings of the above two methods.

Currently, deep learning is widely used in the field of natural language processing, and it is also gradually being applied to text sentiment analysis and classification. Ye [7] proposed a target-specific text sentiment analysis model based on a convolutional neural network, which can effectively realize target-specific text sentiment analysis. Li et al. [8] proposed an XLNet sentiment analysis model incorporating a generalized autoregressive pre-trained language model, which was compared with commonly used sentiment analysis models to verify the model's accuracy. Socher et al. [9] proposed a Recurrent Neural Network (RNN) to realize the selective memory of input information through its recurrent network structure to capture the sequential features of text information. As a type of RNN network, Long Short-Term Memory (LSTM) [10] adds a gating structure based on RNN, turning the model into a nonlinear one, which to a certain degree solves the problem of disappearing or exploding gradient of RNN, and it can memorize the information of earlier time series. Liu et al. [11] proposed a sentiment polarity analysis method based on LSTM and ResNet network, which can obtain higher classification accuracy with guaranteed operational efficiency compared to traditional RNN. Bidirectional Long Short-term Memory (Bi-LSTM) [12] can fully extract the contextual feature information compared to the unidirectional LSTM model. Yang et al. [13] proposed an opinion sentiment analysis model based on ERNIE-BILSTM, using the ERNIE pre-training model to obtain the text to be, and combine it with the bidirectional LSTM to extract the contextual features; the experimental results show that the model predicts better. Bai et al. [14] proposed Temporal Convolutional Networks (TCN) that can parallelize computation and obtain higher-level features with fewer parameters. Gao et al. [15] proposed a text sentiment analysis model based on TCN and bidirectional LSTM, acquiring high-level features by using the causal and dilation convolution features of TCN and extracting bidirectional semantic dependencies of texts by combining the advantages of Bi-LSTM to learn the text contextual information, which overcame the defects of the lack of utilization of contextual information of the common neural network models.

Attention Mechanism [16] in deep learning refers to the process of filtering out the most important information for the target task based on many data samples, thus improving the efficiency and accuracy of task processing. Self-attention Mechanism [17] is a special attention mechanism that pays more attention to the degree of correlation within the data. It is usually used in cooperation with models such as convolutional neural networks, recurrent neural networks, etc. Lin [18] proposed a Chinese microblog sentiment analysis model integrating the attention mechanism and Albert-BiGRU, which significantly improves the prediction effect of the model by giving different attention weights to different words through the design of the attention layer. Liu et al. [19] proposed a sentiment analysis method based on dynamic features of words and self-attention, firstly, dynamic feature encoding of comments based on a pre-training model, and then using the feature reorganization method based on a self-attention mechanism to dynamically integrate the fusion features of words and optimize the weight parameter to reduce the complexity of the algorithm, which is illustrated by experiments to have a better improvement effect.

Most models mentioned above are trained to utilize fixed word vector encoding, either by pre-training a batch of word vectors in advance through techniques such as Word2Vec or by initializing a word vector matrix and then continuously iterating during the training process. Although the Word2Vec method can solve the word-to-word connection issue, it has not yet solved the problem of multiple meanings of one word. Furthermore, Chinese sentences are extremely complicated, with high semantic flexibility and high difficulty in word segmentation, so the above vectorization methods based on word segmentation have certain limitations. Google released the pre-trained model BERT [20], breaking several NLP field task records. After training ultra-large-scale corpus, BERT can obtain a dynamic encoding vector for a word based on the contextual semantic encoding of a text sequence so that during prediction, it can encode dynamically according to different inputs, which solves the limitations of multiple meanings of one word and word segmentation in Chinese sentences.

Consequently, this paper uses the BERT model to obtain sentence-

level semantic information as model input vectors. The proposed model adopts a combined model for sentiment analysis that combines Temporal Convolutional Neural Network (TCN), Bidirectional Long and Short-Term Memory Network (BiLSTM), and Self-Attention mechanism. This model utilizes TCN to obtain higher-level sequence features and then passes them to the BiLSTM layer to fully extract the contextual sentiment features and uses the Self-Attention mechanism to help the model optimize the feature vectors before finally passing them to the fully connected layer with the Sigmoid as the activation function, which achieves a relatively high accuracy of sentiment classification.

3. Model construction

3.1. Word2Vec

Word2Vec is an effective word embedding technique proposed by a research team at Google in 2013 [21]. It captures semantic relationships between words by mapping words into a low-dimensional vector space so that semantically similar words are close to each other in that space. There are two main model architectures for Word2Vec: the continuous bag-of-words model (CBOW) and the skip-gram model (Skip-gram). CBOW predicts the center word through contextual words, while Skip-gram predicts the center word through the contextual words. The core idea of Word2Vec is to utilize contextual information from large amounts of text data to learn distributed representations of words. Compared with the traditional bag-of-words model, Word2Vec captures the frequency information of words and reveals the similarity and association between words. However, Word2Vec generates a fixed vector for each word, which cannot be dynamically adjusted according to the contextual information.

3.2. BERT pre-trained language model

To address the problem of multiple meanings of one word that traditional language models could not solve, we used the BERT pre-trained model that is capable of dynamic semantic coding, the structure of which is shown in Fig. 1:

I1 and I2 denote the input vectors of the model, with a multilayer bidirectional transformer feature extractor in the middle. O1, …, On represent the output vectors of the model. This network model captures more contextual information than the recurrent neural network model, thus obtaining more text features.

BERT is coded based on character level, and the input for each symbol consists of three layers: Position Embeddings, Segment Embeddings, and Token Embeddings. The Token Embeddings layer converts each word in a sentence into a fixed dimensional vector, and the Segment Embeddings represent segment vectors. This layer consists of two kinds of vector representations. If there are two sentences, each token of the first sentence is assigned a value of 0, and each token of the second sentence is assigned a value of 1. If the model input is a single statement, all the tokens of the sentence are 0. Transformers cannot encode the order of the input sequences, and the Position Embeddings layer lets the BERT learn, at each position, a vector to represent the order of the sequence. Position Embeddings represent the position vectors. Finally, the three vectors are summed up to form the input part of the sentiment analysis model, as shown in Fig. 2.

The BERT pre-trained model uses the Transformer to extract features [8]. The Transformer itself consists of multiple units. Each unit includes feed-forward neural networks and self-attention mechanisms, with residual connections designed between layers within the unit. Residual connections are introduced between the layers of each unit. The input data first passes through the self-attention layer of the first sub-layer and then undergoes residual processing and normalization before reaching the second sub-layer. The output is passed to the feedforward neural network of the second sub-layer, which also undergoes residual processing and layer normalization. The linguistic knowledge learned by the BERT and the parameters of the Attention layer are closely related to each other.

3.3. Temporal convolutional neural network

TCN is a neural network model for processing time series and text data. Compared with the traditional RNN, it uses convolutional layers to capture the temporal patterns and features in time series data, and it can process the time series' different time steps in parallel, making it more efficient in the training and prediction process. Its network layer structure is shown in Fig. 3:

In the figure, x_t is the Time Series Data, y_t is the prediction, and d denotes the dilation rate of each convolutional layer. TCN consists of two important operations.

3.3.1. Causal convolution

Causal convolution refers to the fact that an element in the output sequence only relates to the element that precedes it. That is, according to the input sequence $x_1, x_2, \ldots x_{t-1}$ and the current moment input x_t to calculate to get the information of the current moment, the calculation formula is as follows:1

$$P(x) = \prod_{t=1}^{T} p(x_t | x_1, x_2, \ldots x_{t-1}) \tag{1}$$

Causal convolution can remember historical information better than the traditional CNN network, improving the prediction of time series data.

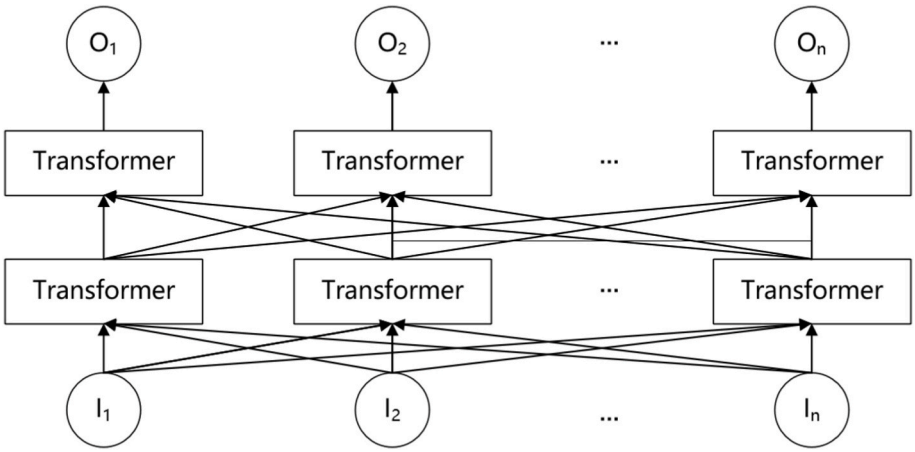


Fig. 1. BERT pre-trained model.

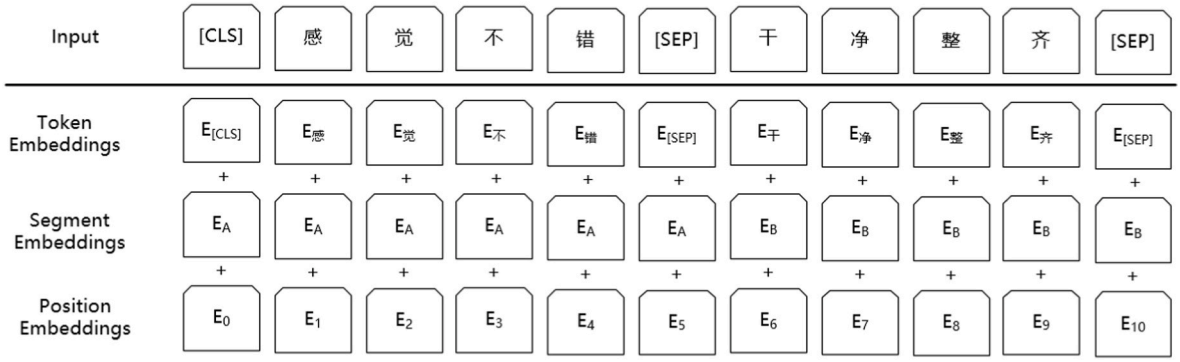


Fig. 2. BERT pre-trained model word vector composition.

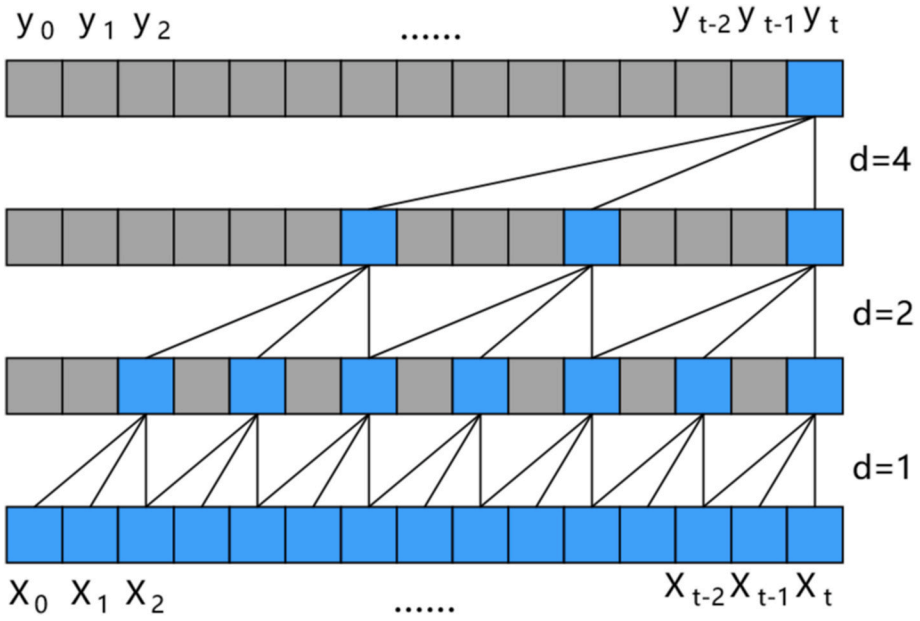


Fig. 3. TCN network structure.

3.3.2. *Dilation convolution*

Dilation convolution is also called atrous convolution. Expanding the receptive field by adding injected voids to the convolution kernel is equivalent to increasing the size of the convolution kernel. Learning text features for as long as possible can be done by increasing the depth of the network layers. However, as the number of layers increases, the problem of gradient vanishing may arise, so residual connections are introduced into the network structure to solve it.

3.4. *BiLSTM algorithm model*

3.4.1. *LSTM model*

The LSTM model is a variant of recurrent neural networks with a core of memory cell units and gating structures that can obtain long-term dependencies in sequences. Memory cells are used to transmit and store information, and gating structures control the addition or forgetting of information. LSTM solves the gradient explosion and gradient vanishing problems that exist in traditional RNN. The LSTM model can be optimized based on Adam's algorithm [22]. The structure of the LSTM gating module is shown in Fig. 4:

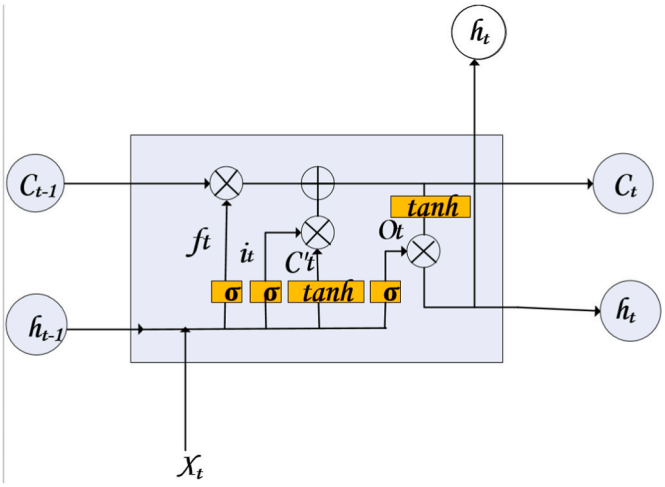


Fig. 4. Structure of LSTM gating module.

The inputs of all three gating include the current input x_t with the hidden state h_{t-1} from the previous step and the output using a sigmoid activation function.

Let i_t, f_t, and o_t represent the values of the input, forgetting, and output gating now, respectively, x_t indicates the input value at moment t, h_{t-1} denotes the final output of the model at the previous time step, W and b stand for the weight matrix and the bias, respectively, σ means the

sigmoid function, and tanh the hyperbolic tangent activation function.

The LSTM network model is computed as follows.

(1) Calculate the forgetting gating, which controls the information to be forgotten, with the following formula:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f\right) \tag{2}$$

(2) Calculate the input gating to control the amount of information added to memory cells with the following formula:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \tag{3}$$

(3) Calculate the output gating with the following formula:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \tag{4}$$

(4) Control the information that needs to be memorized, the formula of candidate memory cell C'$_t$ is as follows:

$$C'_t = \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{5}$$

The formula for the cell state C_t is as follows:

$$C_t = f_t C_{t-1} + i_t C'_t \tag{6}$$

(5) ht is the final output of the model at moment t. The formula is shown in equation (7):

$$y_t = o_t \tanh\left(C_t\right) \tag{7}$$

3.4.2. Bidirectional LSTM model

Temporal information is transmitted unidirectionally from front to back in unidirectional LSTM networks, which fails to consider the contextual temporal information comprehensively. For one certain word of a sentence, the unidirectional model can only obtain the information in front of that word, but in fact, the information before and after it both have a significant effect on the prediction of it, which requires the use of a bi-directional LSTM network. Bidirectional LSTM models can train two LSTMs in opposite directions at the same time, i.e., they can obtain the information before and after a word at the same time and use them for prediction, so generally speaking, bidirectional LSTMs are more effective than unidirectional LSTMs.

The final output of the bidirectional LSTM model is a sum of the outputs of the two unidirectional LSTM networks. Assuming that the forward and backward outputs are $\underline{h_t}$ and $\underline{h_t}$ respectively, the output formula of the bidirectional LSTM model is:

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \tag{8}$$

Both the forward and backward networks of BiLSTM use the same number of neural units in the hidden layer. The specific structure of the BiLSTM model is shown in Fig. 5:

In Fig. 5, $\{h_0 \rightarrow h_1 \rightarrow h_2 \rightarrow \ \ldots \ \rightarrow h_n\}$ are sequences of hidden states generated for a forward LSTM, $\{h_n \rightarrow \ \ldots \ h_2 \rightarrow h_1 \rightarrow h_0\}$ are a sequence of hidden states generated for a backward LSTM.

3.5. Self-attention mechanism

In addition to the need to consider the sequence of contextual information in text sentiment analysis, the importance of sentiment features also has a certain impact on sentiment analysis. The self-attention mechanism is very capable of distinguishing the importance of sentiment features. The attention mechanism operates similarly to the human visual system, which can focus on specific information and capture the important features of relevant information. The self-attention mechanism can better capture the syntactic and semantic features in the same sentence and distinguish the importance of emotional features in it.

The structure diagram of the self-attention mechanism is shown in Fig. 6:

In the self-attention mechanism, the representation vector of each word (token) of the input X is operated with the corresponding weight matrix to obtain three vectors, namely the query vector (Q), the key vector (K), and the value vector (V). To compute the output corresponding to each input, the computational steps are as follows.

(1) Vectors Q and K undergo dot product calculation, i.e., multiply the Query of the current token and the Key vectors of other tokens by the dot product so that the importance degree s of each token would be computed, and the computation formula is as follows:

$$s = QK^T \tag{9}$$

(2) Smooths by SoftMax and multiply it with the value vector V to get the output vector Z, where d_k is the penalty factor. The output vector is computed as follows:

$$Z = \text{softmax}\left(\frac{S}{\sqrt{d_k}}V\right) \tag{10}$$

Performing the same operation on each token would eventually generate a new representation vector for each token, which contains its contextual information. In this paper, based on the BiLSTM module fully extracting contextual sentiment features, the feature vector is optimized by using the self-attention mechanism so that the model pays more

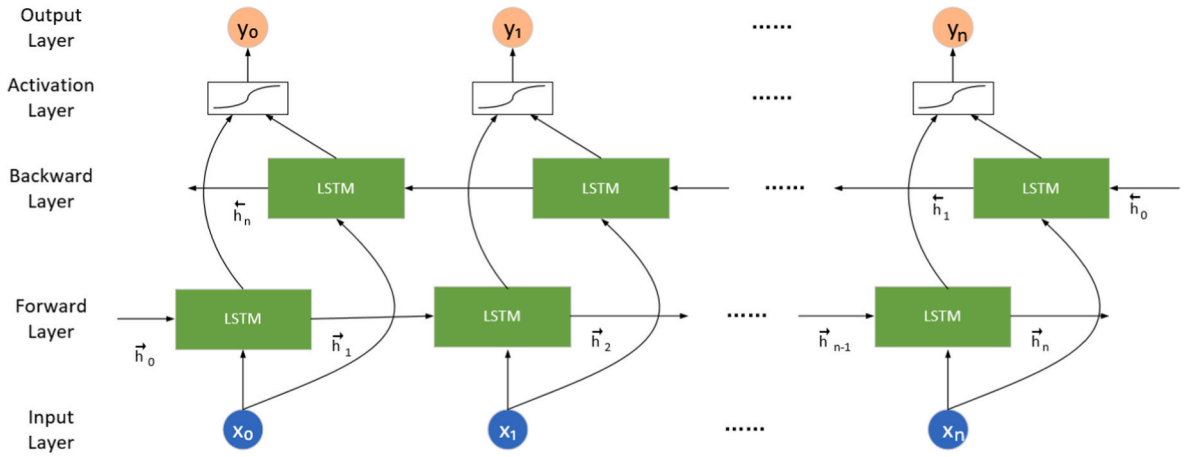


Fig. 5. Structure of BiLSTM model.

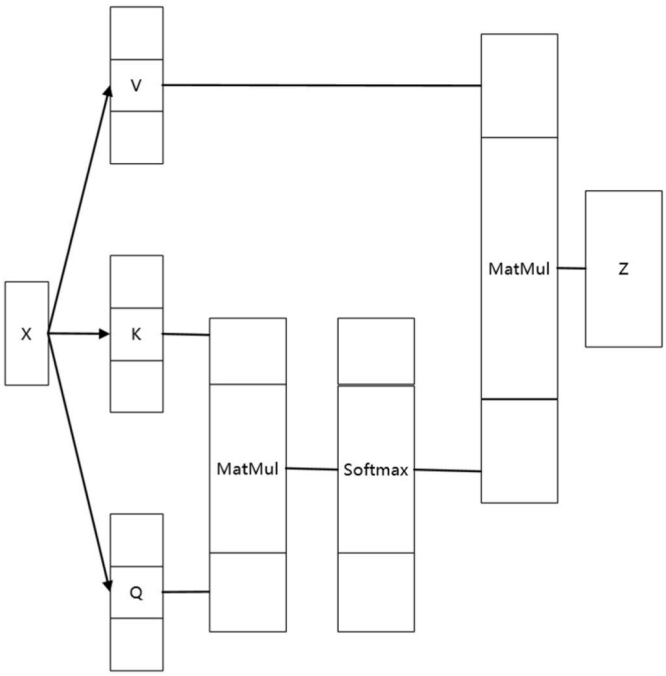


Fig. 6. Structure of the self-attention mechanism.

attention to specific information and enhances the module's ability to extract important features.

3.6. Construction of Chinese sentiment analysis model

This paper proposes a Chinese text sentiment analysis model based on BERT-TCN-BiLSTM-Attention, the structure of which is shown in Fig. 7:

The sentiment analysis model in this paper mainly considers the influence of the comment text's dynamic semantic encoding, the contextual information of the text sequence, and the importance of the sentiment features on the final classification results.

Firstly, in the first layer, the Chinese text based on the hotel review dataset is imported, and the input layer is based on character-level encoding. Each symbol consists of three parts: Token Embeddings, Segment Embeddings, and Position Embeddings, and these vectors of three parts are superimposed to form vectors as input. The second layer is the vector representation layer, where the text phrases from the first layer are dynamically semantically encoded by the Bert pre-trained model, thus finally outputting sentence-level text vectors. The third layer is the TCN network layer, which inputs the text vectors into the network model and expands the receptive field size of the convolution by setting multiple dilations and causal convolution layers to be able to capture the high-level semantic features of the text sequences; the fourth layer is the BiLSTM network model, which takes the output vectors of the TCN as inputs, and combines the two spreading directions of the model through both the forward and backward LSTMs, joints the features learned by the two unidirectional LSTMs to extract the contextual information of the text sequence further fully; the fifth layer is the self-attention mechanism, the vector set of the BiLSTM layer is used as the input of the self-attention mechanism, and the generated weight vector is computed for each word or phrase vector to compute the importance degree of each sentiment feature. Then the output vector of BiLSTM is multiplied by the weight vector to obtain the sentence-level feature vector h* for the final sentiment classification; the last layer completes the text sentiment classification by taking h* as input and obtaining the sentiment classification results through a full connected layer and an output layer with a sigmoid activation function.

4. Experiments and results analysis

4.1. Pre-processing of data samples

The dataset of this paper comes from Dr. Songbo Tan's hotel review dataset ChnSentiCorp-Htl-ba-6000 and ChnSentiCorp-Htl-ba-10000, denoted as C1 and C2, respectively. The ChnSentiCorp-Htl-ba-6000 dataset contains 3,000 records with sentiment classified as positive and 3000 records categorized as negative. ChnSentiCorp-Htl-ba-10000 contains 7000 positive records and 3000 negative records inside. In this paper, the ratio of the training set and the test set is 8:2. To process the data with the word vector model used, the positive and negative review texts are now synthesized into one text, respectively. Part of the text dataset of positive and negative comments is shown in Table 1:

Furthermore, some irrelevant characters in the text file are removed, including English, numbers, and punctuation marks in the text, all of which do not have an impact on the sentiment classification results and are removed in the preprocessing stage.

4.2. Description of model evaluation index

After noise reduction processing and feature extraction, the sample

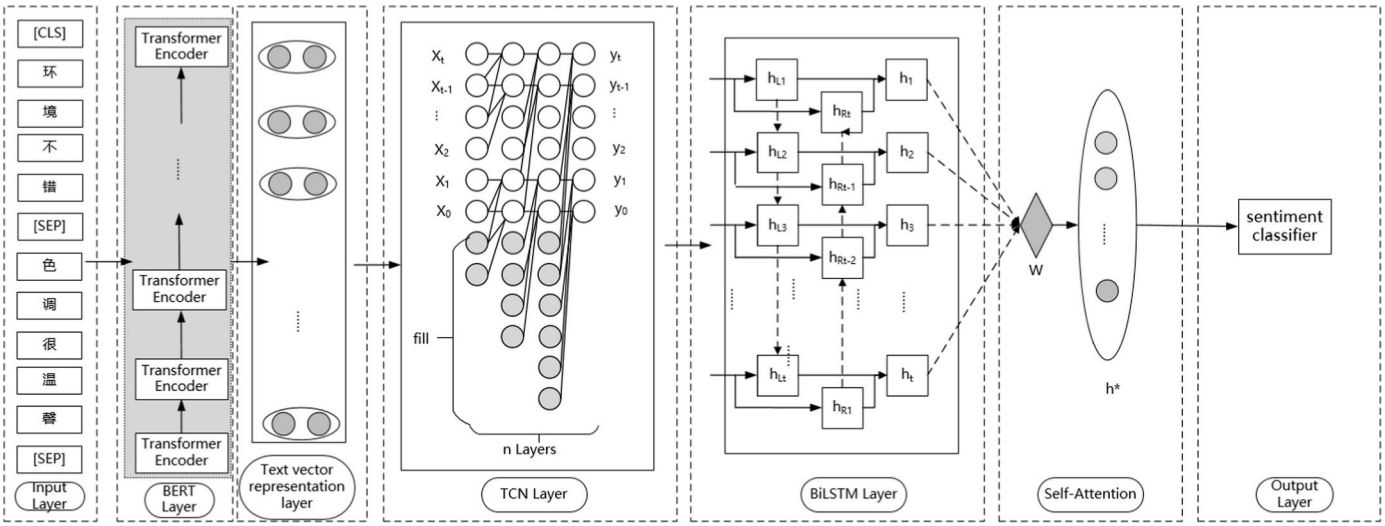


Fig. 7. Structure of the network model proposed in this paper.

Table 1
Part of the sample dataset.

Positive comments	Negative comments
The staff are friendly and professional enough, the rooms are comfortable	Transportation is not very convenient; it is not recommended to choose this hotel
The location of the hotel is at the center of the city, and the rooms are nice	The environment is average. After staying in the hotel, one feels that the price and service do not match.
The rooms are very clean and comfortable, the front desk clerk had a great attitude	The hotel facilities are aged seriously
Location, facilities are good, the price is also okay	It's so bad that it's worse than a hotel with no stars.

data are passed to the model proposed in this paper for model training. The LSTM model is optimized using Adam's algorithm [21], the weights are updated by setting the learning rate, and finally, the model's performance is tested using the test set.

In this paper, four evaluation indicators are used to reflect the model's classification effect: F1 value, accuracy, precision, and recall. These indicators are calculated using TP, TN, FP, and FN.

Among which TP means that both the predicted and true values are true; TN means that both the predicted and true values are false; FP means that the predicted value is true, but the true value is false; and FN means that the predicted value is false but the true value is true.

The specific formulas for each indicator are shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

$$\text{F1} = \frac{2{*}Precision{*}Recall}{Precision + Recall} \tag{14}$$

4.3. Parameter settings

The models used in the experiments and the settings of some key parameters of these models are shown in Table 2. In addition, the model was trained with a batch size of 32, the optimizer used was Adam, and the learning rate was set to Le-5.

4.4. Comparative experiments

Four groups of comparison experiments were conducted in the same experimental setting, corresponding to four experimental models. The first group is the Word2Vec-BiLSTM model, noted as M1. This model first employs the CBOW training model of Word2Vec to generate sentence-level vectors, which are then input to the BiLSTM model. The second group is the BERT-BiLSTM model, denoted as M2. which employs Bert to obtain the dynamic semantic coding vectors of sentences and then inputs them to the BiLSTM model for training tests. The third group is the BERT-TCN-BiLSTM model, denoted as M3. Compared with the second group, a TCN layer is added before BiLSTM to obtain higher-level semantic information. The fourth group is the BERT-TCN-BiLSTM-Attention model proposed in this paper, which is denoted as M4, i.e., a self-attention layer is added on top of the third group to differentiate the importance of sentiment features.

Table 2
Model parameters.

Models	Parameters	Value
BERT	Version	base
	Encoder layers	12
	Number of units per layer	768
	Number of heads	12
	Maximum sentence length	100
	Total parameter size	110M
BiLSTM	Hidden layer units	128
	Dropout rate	0.2
TCN	Dilated causal convolution layers	4
	Convolution kernel size:	7
	Dilation factor	2n
Word2Vec	Training model	CBOW
	Feature vector dimension	100

When the ratio of the training set and validation set is 8:2 and 7:3, respectively, the curves of loss and accuracy change corresponding to the number of iterations for the training set and validation set of the four models M1∼M4 are shown in Figs. 8 and 9.

To evaluate each sentiment categorization model, 1000 records were selected as a test set on the Chinese dataset of hotel reviews, including the number of positive sentiment and negative sentiment categories are 499 and 501, respectively. Based on this dataset, four sets of controlled experiments were conducted, and the confusion matrix was used to show each model's positive and negative sentiment categorization effect. The confusion matrix compares the actual labels with the predicted labels, showing the relationship between the actual and predicted categories of each relationship between the actual and predicted categories of the models, including four parts: the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) categories, as shown in Table 3:

Four groups of controlled experiments were conducted using a Chinese dataset containing hotel reviews. The models were evaluated based on their accuracy, recall, precision, and F1 score on the test dataset. The detailed results of these experiments are presented in Table 4.

To verify the change in the performance of the model proposed in this paper with the increase in the number of iterations, the model is trained using epochs of 30, 50, and 80, respectively and tested based on the C1 dataset. The experimental results are shown in Table 5:

To demonstrate the advantages of the models proposed in this paper more fully, comparative experiments are conducted on the C2 dataset, with the parameter settings of each model kept unchanged. The experimental results are shown in Table 6:

The change curves of the Loss value and accuracy rate of each model on the C2 dataset are shown in Figs. 10 and 11.

4.5. Discussion

From Figs. 8 and 9, it can be seen that the accuracy and loss curves for each model in Fig. 8 are smoother, and the models perform better. So, the ratio 8:2 divides the dataset and trains the models. The trend of the accuracy and loss value curves of the four models in Fig. 8 is basically the same. Still, the minimum loss value on the validation set decreases sequentially, where the minimum loss value of the model proposed in this paper is lower than the other models. The accuracy of the four models, both in the training and validation sets, keeps increasing rapidly in the first five iterations and then fluctuates gently and stabilizes. Among them, the accuracy of the M4 model in the validation set is closer to that of the training set than the other models, indicating that its classification effect is better.

As shown in Table 3, the proposed BERT-TCN-BiLSTM-Attention model has a TP of 467, the highest among all the models, indicating that it performs best in positive emotion classification. For negative emotion categorization, although the model does not have the highest TN value, it has a relatively low FP value and the lowest FN value, indicating that the model performs best in reducing the misclassification of negative emotion as positive emotion (FP) and positive emotion as negative emotion (FN). Therefore, combining the positive and negative classification effects, the BERT-TCN-BiLSTM-Attention model performs the best in the hotel review sentiment analysis task and has the best classification effect.

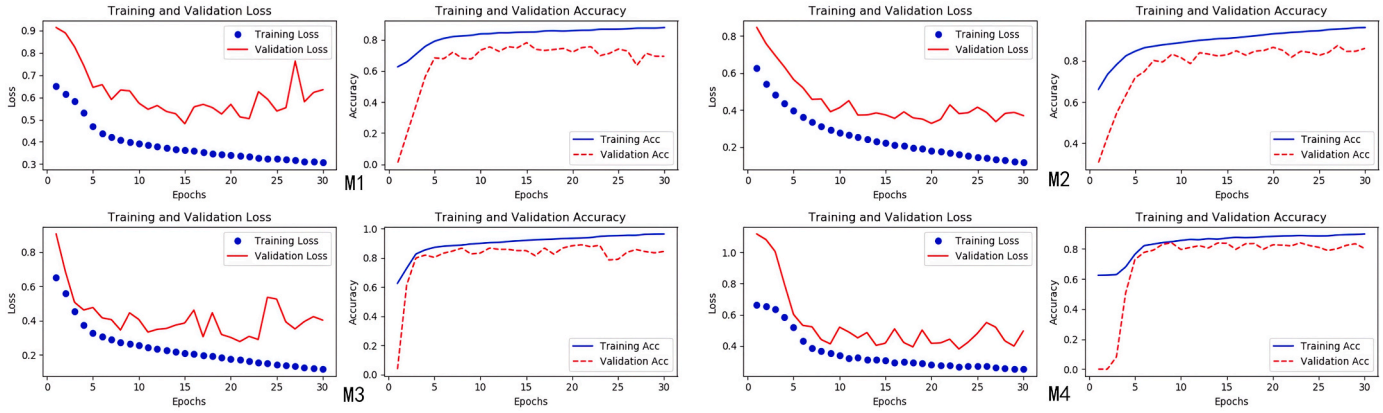


Fig. 8. Effects of loss value vs. accuracy for each model with an 8:2 ratio of training and validation sets.

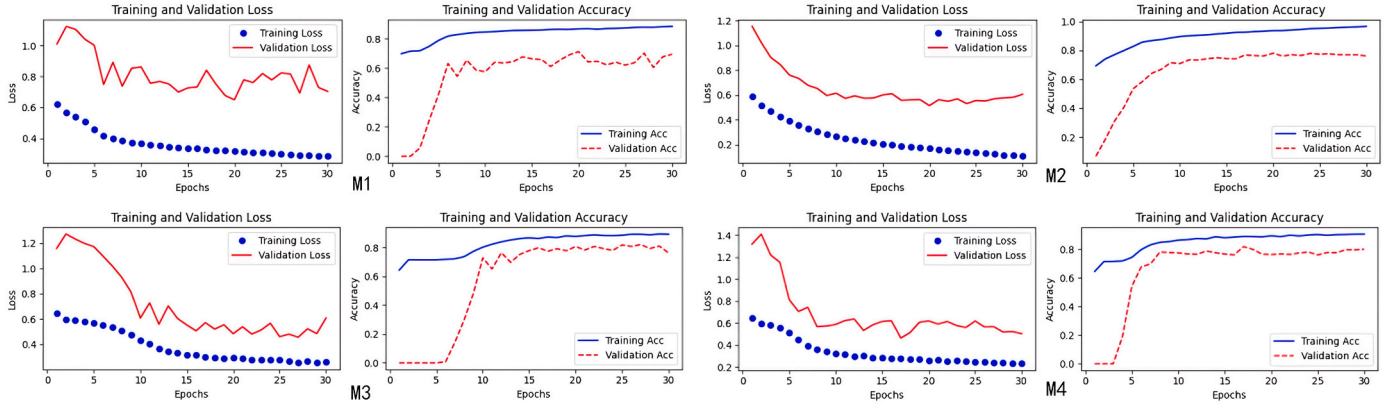


Fig. 9. Effects of loss value vs. accuracy for each model with a 7:3 ratio of training and validation sets.

Table 3
Confusion matrix results for each model.

Model	TP	TN	FP	FN
Word2Vec-BiLSTM	440	373	128	59
BERT-BiLSTM	430	435	65	70
BERT-TCN-BiLSTM	446	432	68	54
BERT-TCN-BiLSTM-Attention	467	427	73	33

Table 4
Comparative experiments results based on the C1 dataset.

Model	Accuracy	Recall	F1	Precision
M1	0.813	0.882	0.825	0.775
M2	0.865	0.860	0.864	0.875
M3	0.875	0.894	0.877	0.868
M4	0.894	0.934	0.898	0.865

Table 5
Experimental results of the proposed model at different number of iterations.

Epochs	Accuracy	Recall	F1 Score	Precision
30	0.894	0.934	0.898	0.874
50	0.890	0.886	0.889	0.893
80	0.889	0.908	0.891	0.875

As seen from the data in Table 4, the M2 model improves the accuracy by 5.2 % and the F1 value by 3.9 % compared with the M1 model. This indicates that the Bert pre-trained model is more capable of

Table 6
Comparative experiment results based on the C2 dataset.

Model	Accuracy	Recall	F1	Precision
M1	0.823	0.898	0.835	0.781
M2	0.850	0.928	0.861	0.803
M3	0.882	0.920	0.887	0.830
M4	0.912	0.933	0.904	0.861

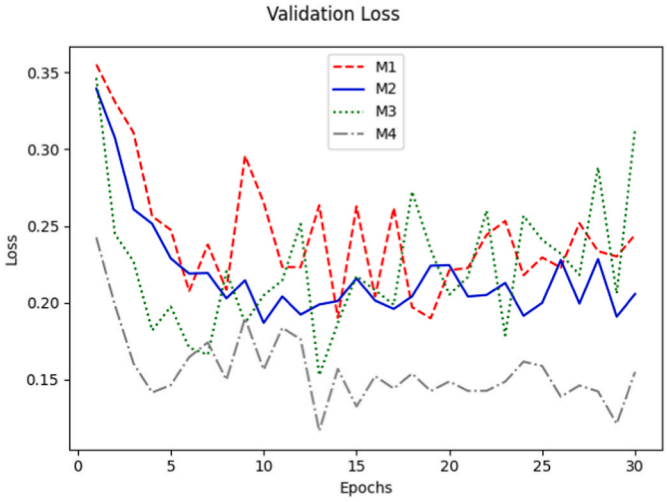


Fig. 10. Changes in loss values for the C2 validation set.

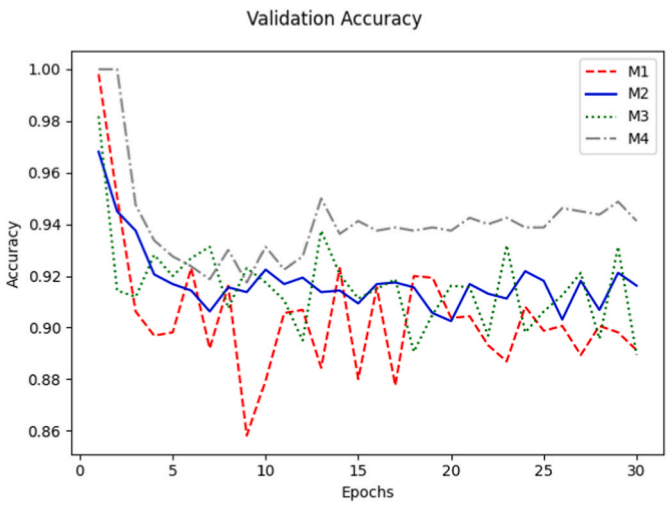


Fig. 11. Changes in accuracy values for the C2 validation set.

extracting the sentiment features of the text than Word2Vec. In particular, the Bert model would judge the actual meanings of the words based on the contextual circumstances, which solves the problem of multiple meanings of one word. So, its accuracy and F1 value of sentiment classification have been significantly improved. M3 has improved its accuracy, recall, and F1 value compared to M2. Among them, the accuracy is improved by 1 %, the recall is increased by 3.4 %, and the F1 value is enhanced by 1.3 %. This indicates that after adding the TCN network, it can capture more valuable textual features., and at the same time, it can make use of its residual link structure to prevent gradient vanishing with too many layers of the network and obtain more information about the sentiment features than BiLSTM alone, which leads to a more accurate classification. The model proposed in this paper is based on M3 with the addition of the attention mechanism. As seen in the table, its accuracy rate reaches 89.4 %, which is further improved by 1.9 % to M3, the recall increases significantly from 89.4 % to 93.4 %, and the F1 value is also enhanced by 2.1 %. This indicates that after adopting the attention mechanism, it can assign different weights to different words, thus better distinguishing the importance of the sentiment features, effectively capturing the local features of the sentence, and improving the model performance.

Table 5 shows that when the number of iterations is 30, the model performs the best in terms of all performance indicators combined.

Table 6 shows that when the dataset increases, the classification effect of each model is higher than that based on the C1 dataset, which indicates that the model can obtain more text semantic features during the training process and thus get better results. At the same time, it can also be clearly seen that the models proposed in this paper have different degrees of improvement in their F1 value, recall value, validation accuracy, and other indicators on this dataset compared with other models.

From Figs. 10 and 11, it can be intuitively seen that the loss value of the proposed model in this paper is always located below the curves of the other models. Its validation correctness curve is always higher than the other curves from the 10th iteration onwards, which proves that the model is more capable in text sentiment analysis. The training and validation on larger datasets make the classification effect of the proposed model in this paper more convincing.

In conclusion, from the experimental results of each model on the test sets, the accuracy of sentiment classification of the model proposed in this paper is comparatively the highest. The text demonstrates that dynamic semantic coding vectors based on BERT are more effective for extracting textual sentiment features compared to Word2Vec, resulting in significantly improved classification accuracy. Additionally, BERT captures contextual information more effectively and places greater emphasis on the key nodes in Time Series Data. By introducing the two-way LSTM model and the self-attention mechanism, the classification results are further enhanced.

5. Conclusion

In this paper, a BERT-TCN-BiLSTM-Attention-based model is proposed to be used for text sentiment analysis of the hotel review dataset in Chinese. Firstly, the data samples are dynamically semantically encoded by Bert to finally form sentence-level text vectors. This effectively avoids the problem of multiple meanings of one word and facilitates the acquisition of effective sentiment features. Then, the text vectors are passed into the TCN network layer to obtain high-level sequence features. Then, they were passed to the BiLSTM model layer to learn the text semantic features from both forward and backward directions to extract the contextual features fully. Finally, the self-attention mechanism and contextual information are used together to distinguish the importance of the sentiment features in the sentence to optimize the feature vectors. Comparative experiments on two hotel review datasets show that the proposed model consistently outperforms the comparison model, which highlights its robustness and practical applicability in sentiment analysis tasks.

CRediT authorship contribution statement

Dianwei Chi: Writing – review & editing, Methodology, Conceptualization. **Tiantian Huang:** Visualization, Software. **Zehao Jia:** Data curation. **Sining Zhang:** Investigation.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

Data will be made available on request.

References

[1] He Y, et al. A deep learning model for sentiment semantic enhancement for Weibo sentiment analysis. J Comput 2017;40(4):773–90. Chinese.
[2] Jiang J. Sentiment analysis of social media texts. Nanjing University of Science and Technology. Chinese; 2017.
[3] Deng C, Yu G. A study on sentiment analysis of movie reviews based on Naive Bayes. Intell Comput Appl 2023;13(2):210–212+217. Chinese.
[4] Chang C. Research on Chinese Weibo sentiment classification technology based on machine learning. Jiangsu University of Science and Technology; 2019. Chinese.
[5] Zhu Y, Ci Q, Yong C. Research on Tibetan microblog sentiment analysis based on SVM algorithm. Comput Simulat 2022;39(8):226–9. Chinese.
[6] Chen Q, Wang Y. Text sentiment analysis based on SVM with improved particle swarm algorithm. J Fuyang Normal Univ (Nat Sci Ed) 2023;40(1):79–86. https://doi.org/10.14096/j.cnki.cn34-1069/n/2096-9341(2023)01-0079-08. Chinese.
[7] Ye H. A target-specific text sentiment analysis model based on convolutional neural network. J Jishou Univ (Nat Sci Ed) 2024;45(1):24–9. https://doi.org/10.13438/j.cnki.jdzk.2024.01.005. Chinese.
[8] Li D, et al. Sentiment analysis of Chinese text based on XLNet. J Yanshan Univ 2022;46(6):547–53. Chinese.
[9] Socher R, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of the 2011 conference on empirical methods in natural language processing; 2011.
[10] Hochreiter S. Long short-term memory. Neural computation. MIT-Press; 1997.

[11] Liu X, Yang B, Yu Y. A method for sentiment polarity analysis based on LSTM and ResNet networks. J Electron Dev 2023;46(6):1629–33. Chinese.
[12] Plank B, Søgaard A, Goldberg Y. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss 2026;1:412–8. https://doi.org/10.18653/v1/P16-1039.
[13] Yang W, Kong K. Sentiment analysis of social network text based on ERNIE-BiLSTM. J China Acad Electron Inf Technol 2023;18(4):321–7. Chinese.
[14] Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling 2018. arXiv preprint arXiv:1803.01271.
[15] Gui X, Gao Z, Li L. Fusion of TCN and BiLSTM + Attention model for text sentiment analysis during the pandemic. J Xi'an Univ Technol 2021;37(1):113–21. https://doi.org/10.19322/j.cnki.issn.1006-4710.2021.01.016. Chinese.
[16] Wang W, Shen J, Jia Y. A review of visual attention detection. J Software 2019;30(2):416–39. https://doi.org/10.13328/j.cnki.jos.005636. Chinese.
[17] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in Neural Information Processing Systems 2017;30:5998–6008.
[18] Lin W. Sentiment analysis of Chinese microblogs incorporating attention mechanism and Albert-BiGRU. J People's Public Security Univ China (Sci Technol) 2023;29(4):52–6. Chinese.
[19] Liu Q, et al. A sentiment analysis method based on word dynamic features and self-attention. J Data Acquis Process 2024;39(1):193–203. https://doi.org/10.16337/j.1004-9037.2024.01.017. Chinese.
[20] Pang L, et al. Text matching as image recognition. In: Proceedings of the AAAI conference on artificial intelligence; 2016.
[21] Mikolov T, et al. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 2013;26.
[22] Kinga D, Adam JB. A method for stochastic optimization. In: International conference on learning representations (ICLR); 2015. San Diego, California.