

ITC 2/52 Information Technology and Control Vol. 52 / No. 2 / 2023 pp. 263-275 DOI 10.5755/j01.itc.52.2.31803	A Multi-Channel Text Sentiment Analysis Model Integrating Pre-training Mechanism	
	Received 2022/07/10	Accepted after revision 2023/03/01
	HOW TO CITE: Liang, S., Jin, J., Du, W., Qu, S. (2023). A Multi-Channel Text Sentiment Analysis Model Integrating Pre-training Mechanism. <i>Information Technology and Control</i> , 52(2), 263-275. https://doi.org/10.5755/j01.itc.52.2.31803	

A Multi-Channel Text Sentiment Analysis Model Integrating Pre-training Mechanism

Shengbin Liang

School of Software, Henan University, Kaifeng, 475004, China;
 Institute for Data Engineering and Science, University of Saint Joseph, Macao, 999078, China;
 e-mail: liangshenbin@henu.edu.cn; liang.shengbin@usj.edu.mo

Jiangyong Jin

School of Software, Henan University, Kaifeng, 475004, China; e-mail: jinjiangyong@vip.henu.edu.cn

Wencai Du

Institute for Data Engineering and Science, University of Saint Joseph, Macao, 999078, China
 e-mail: george.du@usj.edu.mo

Shenming Qu

School of Software, Henan University, Kaifeng, 475004, China; e-mail: qsm@vip.henu.edu.cn

Corresponding author: qsm@vip.henu.edu.cn, george.du@usj.edu.mo

The number of tourist attractions reviews, travel notes and other texts has grown exponentially in the Internet age. Effectively mining users' potential opinions and emotions on tourist attractions, and helping to provide users with better recommendation services, which is of great practical significance. This paper proposes a multi-channel neural network model called Pre-BiLSTM combined with a pre-training mechanism. The model uses a combination of coarse and fine-granularity strategies to extract the features of text information such as reviews and travel notes to improve the performance of text sentiment analysis. First, we construct three channels and use the improved BERT and skip-gram methods with negative sampling to vectorize the word-level and vocabulary-level text, respectively, so as to obtain more abundant textual information. Second, we use the pre-training mechanism of BERT to generate deep bidirectional language representation relationships. Third, the vectors of the three channels are input into the BiLSTM network in parallel to extract global and local features. Finally, the model fuses the text features of the three channels and classifies them using SoftMax classifier. Furthermore, numerical experiments are conducted to demonstrate that Pre-BiLSTM outperforms the baselines by 6.27%, 12.83% and 18.12% in average in terms of accuracy, precision and F1-score.

KEYWORDS: Sentiment Analysis, Pre-training mechanism, Tourism Texts, Multi-channel, BERT, BiLSTM.

1. Introduction

Text sentiment analysis, also known as opinion mining, aims to analyze people's subjective texts and explore the potential emotions, attitudes and opinion tendencies embedded in the text. In the Internet age, it plays an important role in economic and social management and it is widely used in the fields of web opinion analysis, user portrait and intelligent recommendation. Traditional sentiment analysis methods include two common approaches:

Dictionary-based sentiment analysis methods.

This approach need to construct sentiment dictionaries, for example, the more common English dictionary is WordNet [1] and Chinese dictionary is Hownet [17]. This approach does not require training. Cui et al. [4] constructed extended sentiment dictionaries such as adverb dictionaries, network dictionaries and negation dictionaries to classify the sentiment of Weibo topics. The methods have good effects on topic classification, but the dictionary-based sentiment analysis methods have also some disadvantages, such as low accuracy and need for continuous maintenance of dictionaries.

Machine learning-based methods. This approach includes support-vector machine [20], logistic regression [25], Naïve Bayes [31] and random forest [30] etc., however, these methods still have certain problems in feature extraction and efficiency when dealing with high-dimensional data and sparsity. Therefore, deep learning-based sentiment analysis methods and techniques are currently emerging to optimize the accuracy of sentiment analysis.

Ouyang et al. [24] first combined Word2Vec with convolutional neural network (CNN) to solve the problem of feature extraction by using Word2Vec to compute word vectors and taking them as input to CNN. Dos Santos et al. [8] proposed the CharSCNN model, which use information from characters to sentences to solve the sentiment analysis of short texts. Xu et al. [35] proposed an improved word representation method which incorporated sentiment information into the traditional TF-IDF algorithm, generated weighted word vectors, and then input them into BiLSTM model, which can effectively capture contextual information and better represent comment vectors. Munikar et al. [21] introduced BERT into sentiment analysis to address fine-grained problems in senti-

ment analysis tasks. In order to make the sentiment analysis model suitable for both long reviews and short tweets, Basiri et al. [2] proposed the ABCDM model to solve this problem. Sun et al. [29] showed the ERNIE 2.0 text pre-training model, which achieved excellent results in all 16 types of tasks in Chinese natural language processing. Wu et al. [34] introduced a new model, SC-ABiLSTM, which can quickly identify sentimental words from a large number of words and classify them into several different types, improving the accuracy of sentiment analysis on microblogging datasets.

The above models have made great progress in feature representation, bidirectional learning and fine-tuning parameters, but the length of tourist attractions review texts is very different. Furthermore, the pre-training of models often train the model based on the co-occurrence of words and sentences. Therefore, the pre-training mechanism introduced BiLSTM can better obtain context-related bidirectional features, and has the ability to process sentences or paragraphs to solve the problem of long-distance dependencies. Meanwhile, we use a multi-channel mechanism to implement compute in parallel and multi-task learning. As we all know, tourism industry is an experiential service. Today, with the development of social media, the evaluation information of tourist attractions has an important influence on the intention choices of tourists. Therefore, sentiment analysis of tourist attractions reviews, travel notes and other data play an important role in analyzing tourists' satisfaction with tourist attractions, helping tourists choose high-quality tourist attractions, and improving the service quality of tourist attractions. However, the traditional methods such as questionnaires to analyze the satisfaction of tourist attractions are limited by the small amount of data and lack of comprehensiveness, due to people's inherent emotional cognition [7] and the passiveness of information feedback, the results have a large deviation from the actual state. On the contrary, comments on social media are a kind of active feedback, which is less affected by subjective influence; however, there are some problems such as too concentrated polar comments and short texts. As an important application in the field of natural lan-

guage processing (NLP), text sentiment analysis analyzes people's opinions, attitudes and emotions towards tourist attractions, which has reference value for potential tourists to choose tourist attractions.

In summary, our contributions are as follows:

- **Propose a multi-channel architecture.** A multi-channel text processing architecture is proposed to extract text feature with different granularity, and it has parallel computing capability, which can extract bidirectional word-level and character-level feature.
- **Adopt pre-training mechanism.** In order to fully mine textual context information and solve the problem of long-distance dependence of texts, a pre-training mechanism is introduced to take into account both global and local features to improve the performance.
- Several experiments are conducted on three datasets, and our model has better performance in terms of accuracy, precision and F1-score compared with baselines.

The structure of this paper is as follows: the second section introduces the related work and background. The third section gives the problem definition and then focuses on the model definition. The fourth section gives the experimental results, and, in the final section, the conclusion remarks are presented.

2. Related Work

The current deep learning technology has achieved breakthrough results in the field of text sentiment analysis. Generally, the basic process of text sentiment analysis based on deep learning is the following: text vectorization, model construction, model training, and model evaluation.

2.1. BERT

Essentially, data should be preprocessed to meet the requirements of the model. Traditional data preprocessing methods such as TF-IDF, and deep learning-based methods such as word embedding. Word embedding methods include Word2Vec, Item2Vec, GloVe and BERT.

Karimi et al. [11] propose a novel architecture called BERT Adversarial Training (BAT) to utilize adver-

sarial training for the two major tasks of Aspect Extraction and Aspect Sentiment Classification in sentiment analysis. Yang et al. [37] propose a CM-BERT model that applies BERT to multimodal sentiment analysis, which combines information from text and audio modalities to dynamically adjust the word weights to fine-tune the pre-trained BERT model, which has achieved better accuracy. Jain et al. [10] propose a BERT-DCNN model to address the problem that traditional CNNs are insufficient to understand the contextual semantics in long-term sequences used for sentiment classification. Yan et al. [36] argue that pre-trained models such as BERT lack domain-specific knowledge, and therefore propose the SAKG-BERT model to address domain-specific sentiment analysis by combining it with knowledge graphs. VGCN-BERT [18] takes advantage of graph convolutional neural networks to solve the problem of insufficient global information captured by BERT for vocabulary, which interacts local and global information through different layers of BERT so that they can interact and jointly build the final sentiment representation.

However, the above research methods essentially rely on the BERT model to implement text encoding and solve problems such as feature extraction of long sequence texts. Actually, BERT's pre-training task MLM enables the encoding of sequences with context, but as the same time makes it difficult to adapt to generative tasks due to the mismatch between the data in the pre-training process and the fine-tuned data.

2.2. LSTM

A big challenge in sentiment analysis is understanding the context of long texts at the paragraph level or even at the document level. CoLSTM [3] is a neural network architecture comprised of both CNN and LSTM to predict the sentiment of customer reviews. Li et al. [16] propose a dual-channel CNN-LSTM model with integrated lexicon for problems caused by text order, sentence length, and complex logical changes in sentiment analysis, which can provide more refined sentiment analysis. Rehman et al. [26] extract local features using CNN, capture long-term dependencies using LSTM, and combine these features into a Hybrid CNN-LSTM model, which achieved better performance in sentiment analysis. Kiran et al. [12]

address the inability of traditional sentiment analysis to capture context and semantics, the OSLCFit model is proposed, which uses migration learning to fine-tune the language model embedding for a specific task of polarity classification so that it can capture contextual semantic information. Ni et al. [23] construct a sentence sentiment analysis system based on the GloVe word vector model and recurrent neural networks, and proposed a neural network model combining LSTM and GRU in order to overcome the shortcomings of recurrent neural networks that cannot learn long-term information about text. SAB-LSTM [13] uses extension of BiLSTM with additional layers to process long text of social media posting, news articles. Some other methods such as GRU, another variant of LSTM to solve specific problems, SLCABG [38] uses CNN to extract the important features in the input matrix, and then uses BiGRU to consider the order information of the input text to extract text context features.

People started to adopt LSTM or its variant to solve the problems of long-distance dependencies and contextual relations. In contrast, our work finds it important to address textual features of different granularities.

3. Methodology

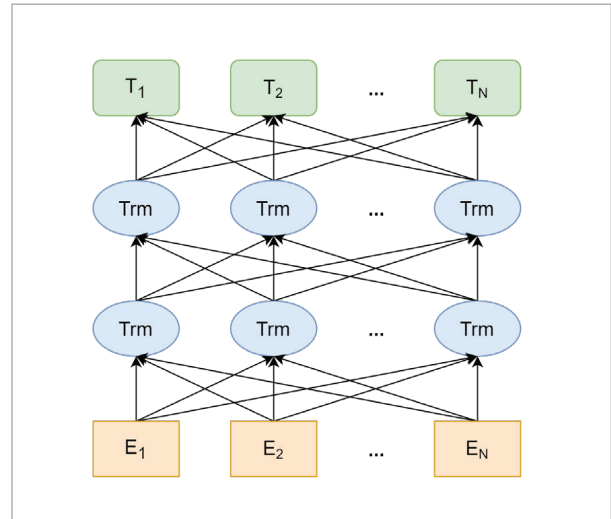
According to the pain point of text sentiment analysis in the tourism field, we propose a multi-channel text sentiment analysis model integrating pre-train mechanism, Pre-BiLSTM. The Pre-BiLSTM model has a multi-channel architecture, and each channel uses different strategies to process text information of different granularity. The three channels use BERT, word embedding and char embedding techniques to extract the sentiment information of comments, and then serve as the input of the BiLSTM neural network to further extract textual contextual information, so that we can more accurately analyze the emotional tendencies of tourists.

3.1. Background

Google AI Language Lab [6] proposed the BERT algorithm that gave a breathtaking performance in 11 downstream tasks in NLP. The structure of BERT adopts multi-layer bidirectional transformer encoder [33], and uses a self-attention mechanism for learn-

ing, so that it can learn the complex relationship of text and the dependency between terms. Its framework is shown in Figure 1.

Figure 1
BERT Framework

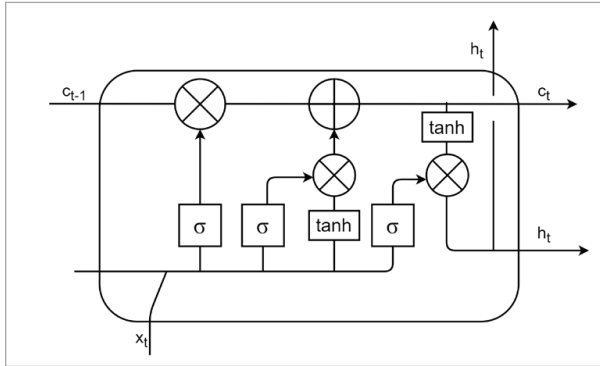


BERT employs two major unsupervised learning tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP), with the aim of minimizing the combined loss function of these two strategies. Among them, MLM replaces the words in each sentence with random probability, and replaces the token in each training sequence with mask token ([MASK]) with 15% probability before feeding the sentence into BERT, which is different from other word embedding methods, such as Word2Vec [19]. NSP focuses on learning the relationship between pairs of sentences, which makes BERT better for various downstream tasks such as sentiment analysis, question and answer, inference, sentence-topic relations, and other NLP tasks.

LSTM was proposed by Hochreiter et al. [9] in 1997. It is used to improve the traditional recurrent neural network (RNN) model. Similar to the principle of RNN, it learns sequence data by repeating the chain form of neural network modules. However, the traditional RNN structure fails to solve the problem of gradient disappearance and explosion, and it cannot handle the problem of long-term dependencies. Therefore, the LSTM architecture provides a series of repeating modules for each time step in a standard

RNN. these modules are called cells. At each time step, it is transmitted between the units of the hidden layer, the “gate” is used to control the discarding or adding of information, so as to achieve the function of forgetting or memory. The LSTM cell structure is shown as Figure 2.

Figure 2
LSTM cell structure



At the time step t , it is assumed that the input and output vectors of LSTM hidden layer are x_t and h_t , respectively, and the memory unit is c_t . The input gate and tanh function which represents a hyperbolic tangent function cooperate to control the addition of new information, and its value is computed as Equation (1),

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i), \tag{1}$$

where W_i is the weight matrix of the input gate, b_i is the bias of the input gate.

The forget gate is a key component of LSTM, which controls what information is retained or forgotten, and avoids the problem of gradient disappearance or explosion caused by gradient over time in some way. It is the output h_{t-1} of the previous unit and the input x_t of current unit as the input of the SoftMax function, and the output value $f_t \in (0,1)$ is assigned to the state c_{t-1} of the current cell, as shown in Equation (2).

$$\begin{aligned} f_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\ q_t &= \tanh(W_q * [h_{t-1}, x_t] + b_q) \\ c_t &= f_t * c_{t-1} + i_t * q_t \end{aligned} \tag{2}$$

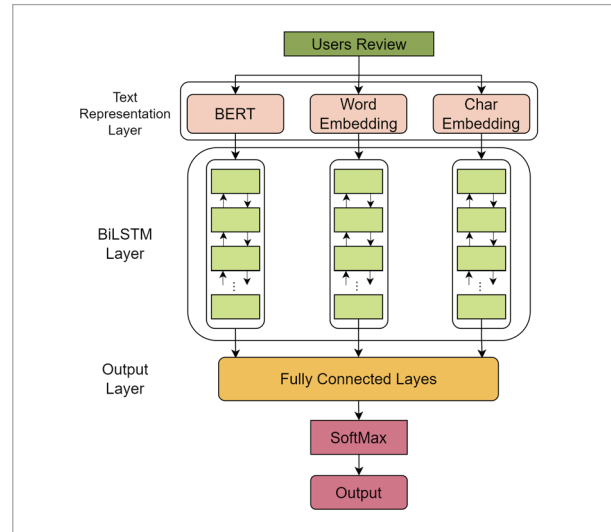
The output gate is based on the cell state, that is, the part of the memory unit will be output at time step t , the value of the output gate is o_t , and the output at time t is h_t , as shown in Equation (3):

$$\begin{aligned} o_t &= \sigma(W_o * [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(c_t) \end{aligned} \tag{3}$$

3.2. Pre-BiLSTM Structure

The text sentiment analysis model proposed in this paper includes four layers, and the overall structure is shown in Figure 3.

Figure 3
Pre-BiLSTM Framework



Channel layer. The channel layer is one of the important innovations of the model proposed in this paper. In order to be comprehensive, accurate and highlight the deep sentiment tendencies embedded in the text, our model constructs three channels in the channel layer to pre-train the text.

Feature extraction layer. This layer uses BiLSTM network to extract features from the vectors processed by the three channels, BiLSTM can address the contextual relationship of the context and highlight key features.

Fully connected layer. This layer performs connected fusion of the features of each of the three channels processed by BiLSTM.

Output layer. This layer uses SoftMax classifier to classify the features fused in the fully connected layer and finally outputs the sentiment analysis results.

The steps of the algorithm are as follows.

Step 1: Data pre-processing, eliminating invalid data from the original data.

Step 2: Split the dataset into training set, validation set and test set.

Step 3: Embedding the dataset separately to obtain word vectors with different granularity.

Step 4: Send word vectors of different granularities into BiLSTM, respectively.

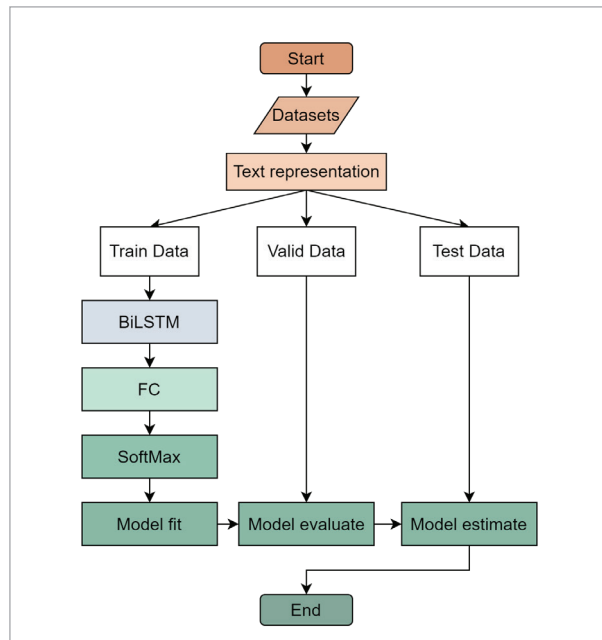
Step 5: Perform a fully connected operation on the outputs of different channels.

Step 6: Use SoftMax classifier to get the final results.

The specific process is shown in Figure 4.

Figure 4

Pre-BiLSTM stack flow



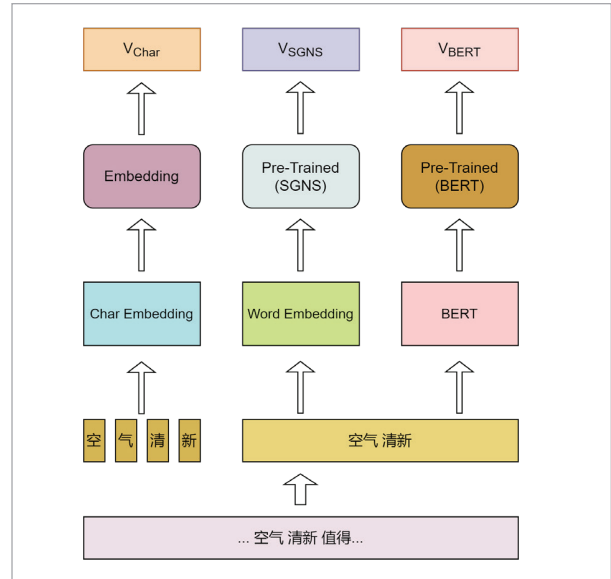
3.3. Channel Layer

The tourist attractions comments published by tourists on tourism website often include multiple information such as text and emoticons, and have the characteristics of variable length and unstructured text. In order to better extract features with complex-

ity, we adopt a three-channel approach that is BERT channel, char channel and word channel. The three channels use BERT, char embedding and word embedding methods to construct text information vectors, respectively. The specific structure of the three channels is shown in Figure 5.

Figure 5

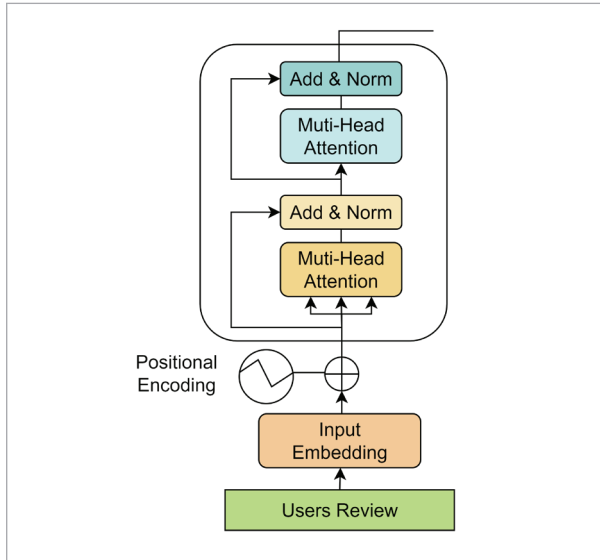
Text representation



BERT Embedding: We first use BERT to extract the text feature information, because the training process of BERT is extremely complex, requires high hardware configuration, long training time and very expensive, so this experiment uses the open source pre-training model by Xunfei Joint Laboratory of HUST for Chinese NLP tasks [5]. In this experiment, we use the pre-training model open-sourced by Google for the English NLP task [32], which has 12 layers of transformer, 768 hidden units and 12 Multi-Attention, and we adapt it for the proposed Chinese dataset with the same parameters. The structure of the single-layer transformer network of BERT is shown in Figure 6. It contains four parts, namely, input layer, attention mechanism, residual connectivity and layer normalization, and feedforward neural network.

The text input layer convert the comment text information into location information for BERT recognition in the form of location encoding. Thus, the features of the text comments are obtained.

Figure 6
Transformer Network



$$X = E(X) + PE, X \in \mathbb{R}^{\text{batch_size} \times \text{seq.len} \times \text{embed_dim}}, \quad (4)$$

where $E(x)$ denotes the word embedding of the corresponding word, PE is the position encoding derived from Equation (5), batch_size denotes the number of input texts, denotes the text length, and denotes the dimension of each word embedding.

$$\begin{aligned} PE_{(\text{pos}, 2i)} &= \sin\left(\frac{\text{pos}}{1000^{2i/d_{\text{model}}}}\right) \\ PE_{(\text{pos}, 2i+1)} &= \cos\left(\frac{\text{pos}}{1000^{2i/d_{\text{model}}}}\right) \end{aligned}, \quad (5)$$

where pos denotes the position of the vocabulary in the text, $2i$ denotes an even position, and $2i+1$ is an odd position.

Next, X enters the attention mechanism module, which ensures that each word vector contains all the word vector information in the comment text.

$$\begin{aligned} Q &= \text{Linear}(X) = XgW_Q \\ K &= \text{Linear}(X) = XgW_K \\ V &= \text{Linear}(X) = XgW_V \end{aligned} \quad (6)$$

$$X_{\text{attention}} = \text{SelfAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (7)$$

where $\text{Linear}(X)$ is the linear mapping and X is the comment text vector. W_Q, W_K, W_V is the weight. $\sqrt{d_k}$ is the attention mechanism transformed into a normalized normal distribution, and $\text{softmax}(\cdot)$ is the sum of the attention weights of each word in the comment text normalized to the attention weights of other words.

After that, it goes to residual concatenation and cascade normalization.

$$X_{\text{attention}} = X_{\text{embedding}} + \text{SelfAttention}(Q, K, V) \quad (8)$$

$$X_{\text{attention}} = \text{LayerNorm}(X_{\text{attention}}) = \alpha \odot \frac{X_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} + \beta, \quad (9)$$

where μ_i is the mean of the text matrix rows, σ_i^2 is the variance of the text matrix rows, \odot is the matrix element multiplication, α and β are the model training parameters, and ε is used to ensure that the denominator is not zero.

Finally, through a feedforward neural network activated by two-layer linear mapping and activation function ReLU, comment information is processed by hidden sequences and calculated as follows.

$$V_{\text{BERT}} = \text{ReLU}(\text{Linear}(\text{Linear}(X_{\text{attention}}))). \quad (10)$$

Word Embedding: To better take advantage of the large amount of repeated data, we use an implicit matrix decomposition model SGNS [14]. This is an unsupervised learning model. It differs from the traditional Word2Vec (CBOW and Skip-gram) in that it introduces negative sampling, using one positive sample and several negative samples in each training, so that the positive samples have the largest possible prediction probability and the negative samples have the smallest possible prediction probability, and the introduction of negative samples saves computing power to achieve the purpose of accelerated training.

In this text representation layer, for Chinese comment text information, we first use the Jieba word splitting tool [28]. For the non-normalized comment information, firstly, we perform the noise reduction process, and then serialize the processed text, finally, we combine the pre-training model SGNS with 299-dimensional Word2Vec embedding to obtain a text matrix with 299-dimensional word vectors V_{SGNS} .

$$S = [w_1, w_2, w_3, \dots, w_m] \quad (11)$$

$$V_{SGNS} = [S_1, S_2, \dots, S_n]^T, \quad (12)$$

where w_m denotes a 299-dimensional word vector corresponding to a word, and S_n denotes a text comment message.

Char Embedding: Due to the complexity of Chinese characters, one character may have rich textual information, so we use character-level features to extract sentiment information V_{Char} , to help us uncover more hidden sentiment. This method is more helpful for extracting ambiguous sentiment information.

3.4. Feature Extraction Layer

The feature extraction layer uses a BiLSTM neural network to capture the contextual information of each channel. In this paper, BiLSTM receives input vector V_{BERT} , V_{SGNS} and V_{Char} , and connecting the hidden representations from LSTM with forward propagation \vec{h}_t and backward propagation \overleftarrow{h}_t , then obtains the hidden representation h_t at time t . Finally, the output of BiLSTM is obtained as O , which can be expressed as.

$$h_t = [\vec{h}_t \parallel \overleftarrow{h}_t] \quad (13)$$

$$O = \sum_{t=1}^z h_t, \quad (14)$$

where w_m denotes the forward and backward propagation of the connected LSTM.

3.5. Fully Connected Layer

The fully-connected layer is responsible for fusing the results of the feature extraction layer for the three channels. The results of the three channels computed by BiLSTM are represented as O_B , O_w and O_c , respectively, they are fused using Equation (15) and output as H .

$$H = [O_B \oplus O_w \oplus O_c]. \quad (15)$$

3.6. Output Layer

This layer takes the features fused in the fully connected layer and calculates the sentiment classification results using the SoftMax classifier, as shown in

Equation (16). The classification results of this model, that is, the values of y include three possible values: positive, negative and neutral.

$$y = \text{Softmax}(\text{dense}(H)). \quad (16)$$

4. Experiments and Results

4.1. Datasets

There are three experimental datasets in this paper, one of which the data is mainly collected from ctrip.com, we call it Tourism dataset; the other dataset is the Amazon Reviews Dataset provided by DataStock, the last dataset uses the dianping.com dataset open sourced by Peking University Open Research Platform [15]. The statistical information of each dataset is shown in Table 1.

Table 1
Statistics of datasets

Dataset	Tourism	Amazon	Dianping
Items	313 328	53 048	325 040
Positive	216 798	42 916	249 204
Negative	46 476	5 985	20 118
Ambiguity	50 054	4 147	55718

The experimental hyperparameters of the Pre-BiLSTM model proposed in this paper are shown in Table 2.

Table 2
Hyperparameter settings

Experimental parameters	Value
Word embedding dim	299
maxlen	50
Batch size	32
Epoch	10
Max features	89993

4.2. Baselines and Evaluation Metrics

Evaluation: For all experiments, we evaluate our model and baselines in terms of accuracy, precision and F1-score. Among them, accuracy represents the overall judgment ability of sentiment analysis, the ratio of the number of correctly predicted samples to the total number of predicted samples; precision represents the ratio of the number of positive samples obtained by sentiment analysis to the total predicted positive samples; recall refers to the proportion of the total number of positive samples obtained from sentiment analysis; F1-score is an indicator that combines precision and recall, and is calculated as follows.

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + FP + TN + FN} \\
 \text{Precision} &= \frac{TP}{TP + FP} \\
 \text{Recall} &= \frac{TP}{TP + FN} \\
 \text{F1-Score} &= \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}
 \end{aligned} \tag{17}$$

Note: TP stands for true example, FN stands for false counterexample, FP stands for false positive example, and TN stands for true counterexample.

Baselines: We compare the Pre-BiLSTM model with a variety of models, including the BERT-based deep neural network method SLCABG [38], the deep neural network method Co-LSTM [3], SC-ABiLSTM [34], ABCDM [2] and ERNIE [29]. To make the comparison results objective, we use the same parameters and training scheme as Pre-BiLSTM for BERT and SLCABG. At the same time, we do some ablation experiments, including basic BERT [27] sentiment analysis, BiLSTM model using word embedding and Dual-BiLSTM model using word embedding and character embedding.

- **Co-LSTM:** It is a hybrid method that combines two deep learning structures, CNN and LSTM, for sentiment analysis of comments published in different domains.
- **SLCABG:** It is based on sentiment lexicon and combined with CNN and Bi-GRU. It combines the points of sentiment lexicon and deep learning techniques, and uses BERT as a pre-training word vector, which shows good results in the sentiment analysis of comments in e-commerce websites.

- **SC-ABiLSTM:** the model proposes a sentiment classification method for large scale microblog text based on the attention mechanism and the bidirectional long short-term memory network.
- **ABCDM:** the model uses two independent bidirectional LSTM and GRU layers to extract past and future contextual information, and uses attention mechanism to emphasize different words. In order to solve low feature dimension and extract local features in the same location, it also uses convolution and pooling mechanism.
- **ERNIE:** in order to extract lexical, syntactic and semantic information from the training corpus, Baidu team proposed a continuous pretraining framework called ERNIE 2.0. It builds pretraining tasks step by step and then learns the pretraining model on these constructed tasks through continuous multi-task learning. Based on this framework, several tasks were constructed and ERNIE 2.0 models were trained to capture lexical, syntactic and semantic aspects of information in training data.

4.3. Results

We reproduce the code of the baseline models and train these models on the Tourism, Amazon and Dianping datasets. Tables 3-5 show the performance of our model and baselines in terms of accuracy, precision and F1-score on different datasets. The graphical representations of the experiments are show in Figures 7-9, respectively.

Table 3

Experimental Results on Tourism Dataset

Models	Accuracy	Precision	F1-score
Co-LSTM	0.7706	0.6603	0.6253
SLCABG	0.7205	0.648	0.6354
SC-ABiLSTM	0.8072	0.7182	0.7076
ABCDM	0.7748	0.6359	0.626
ERNIE	0.8065	0.6934	0.7018
BERT	0.8474	0.7889	0.7702
BiLSTM	0.7431	0.5924	0.5002
Dual-BiLSTM	0.7737	0.6297	0.5459
Pre-BiLSTM	0.8654	0.8152	0.7989

Table 4
Experimental Results on Amazon Dataset

Models	Accuracy	Precision	F1-score
Co-LSTM	0.8451	0.4982	0.4841
SLCABG	0.81	0.7356	0.7554
SC-ABiLSTM	0.9061	0.7704	0.7571
ABCDM	0.9001	0.7416	0.7336
ERNIE	0.908	0.7529	0.7575
BERT	0.8978	0.7594	0.7388
BiLSTM	0.8628	0.6194	0.5317
Dual-BiLSTM	0.8535	0.6581	0.5067
Pre-BiLSTM	0.9118	0.7949	0.762

Table 5
Experimental Results on Dianping Dataset

Models	Accuracy	Precision	F1-score
Co-LSTM	0.7796	0.5714	0.4546
SLCABG	0.7667	0.5879	0.6654
SC-ABiLSTM	0.7887	0.6085	0.516
ABCDM	0.7712	0.6144	0.614
ERNIE	0.8128	0.6488	0.6697
BERT	0.8006	0.6645	0.6343
BiLSTM	0.7736	0.5654	0.3995
Dual-BiLSTM	0.7796	0.6714	0.4047
Pre-BiLSTM	0.8168	0.6812	0.6374

A multi-channel sentiment analysis model can certainly improve Accuracy, but different lengths of datasets have different effects on sentiment analysis results, for example, different length of text sequences have more obvious effects on sentiment analysis. Therefore, when using multi-channel sentiment analysis, pre-analysis of the dataset is particularly important; at the same time, different pre-trained bag-of-words models are selected for different languages to ensure the generalizability of the sentiment analysis model.

Figure 7
Experimental Results on Tourism Dataset

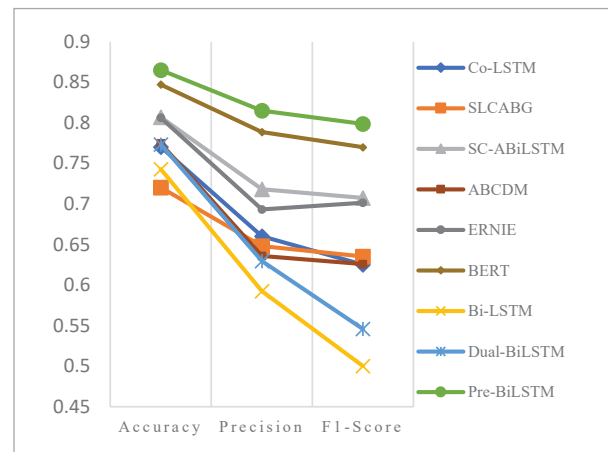


Figure 8
Experimental Results on Amazon Dataset

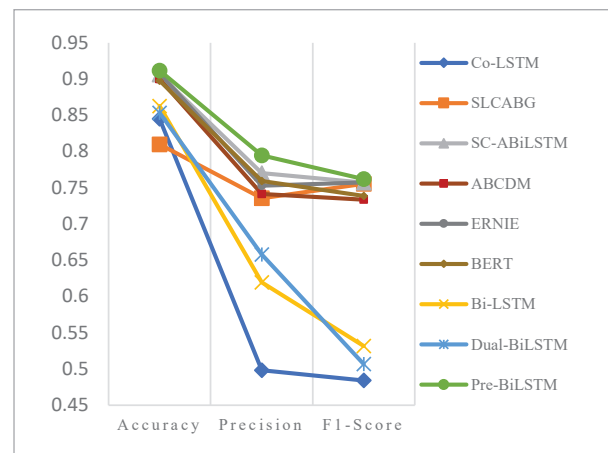
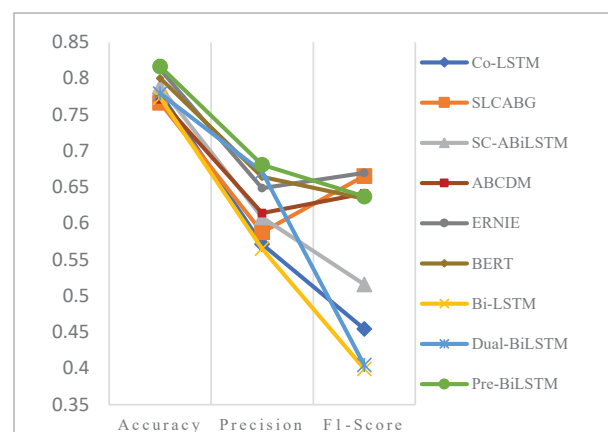


Figure 9
Experimental Results on Dianping Dataset



5. Conclusion and Future Work

Tourists have a higher requirement to obtain the overall sentiment tendency of users towards tourist attractions from the massive reviews and travelogues. In this paper, an improved pre-training sentiment analysis model is proposed, in feature extraction and mining, the multi-channel techniques are used to invoke BERT, word embedding and char embedding with different granularity of text information for feature extraction, while BiLSTM is used in the sentiment analysis task to compute bidirectional analysis of global and local sentiment features, focusing on their contextual information. The experimental results of the Pre-BiLSTM model on three datasets show that the performance is improved by 6.27%, 12.83% and 18.12% on average in terms of accuracy, precision and F1-score compared with baselines.

References

- Banerjee, S., Pedersen, T. An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. Proceedings of the International Conference on Computational Linguistics & Intelligent Text Processing, February 2002, 136-145. https://doi.org/10.1007/3-540-45715-1_11
- Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., Acharya, U. R. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for Sentiment Analysis. Future Generation Computer Systems, 2021, 115, 279-294. <https://doi.org/10.1016/j.future.2020.08.005>
- Behera, R. K., Jena, M., Rath, S. K., Misra, S. Co-LSTM: Convolutional LSTM Model for Sentiment Analysis in Social Big Data. Information Processing & Management. 2021, 58(1), 102435. <https://doi.org/10.1016/j.ipm.2020.102435>
- Cui, A., Zhang, H., Liu, Y., Zhang, M., Ma, S. Lexicon-based Sentiment Analysis on Topical Chinese Microblog Messages. Semantic Web and Web Science, Springer, 2013, 333-344. https://doi.org/10.1007/978-1-4614-6880-6_29
- Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z. Pre-training with Whole Word Masking for Chinese BERT. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29, 3504-14. <https://doi.org/10.1109/TASLP.2021.3124365>
- Devlin, J., Chang, M., Lee, K., Toutanova, K. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, 2018. <https://doi.org/10.48550/arXiv.1810.04805>
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., Megerdooian, K., McMahon, M. T., Tivnan, B. F., Danforth, C. M. Human Language Reveals a Universal Positivity Bias. Psychological and Cognitive Sciences, 2015, 112(8), 2389-2394. <https://doi.org/10.1073/pnas.1411678112>
- Dos Santos, C. N., Gatti, M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. Proceedings of the Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, August, 2014, 69-78.
- Hochreiter, S., Schmidhuber, J. Long Short-Term Memory, Neural Computation. Neural Computation, 1997, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jain, P. K., Quamer, W., Saravanan, V., Pamula, R. Employing BERT-DCNN with Sentic Knowledge Base for Social Media Sentiment Analysis. Journal of Ambient Intelligence and Humanized Computing, 2022, 1-13. <https://doi.org/10.1007/s12652-022-03698-z>
- Karimi, A., Rossi, L., Prati, A. Adversarial Training for Aspect-based Sentiment Analysis with BERT. Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), January, 2021, 8797-8803. <https://doi.org/10.1109/ICPR48806.2021.9412167>

12. Kiran, R., Kumar, P., Bhasker, B. Oslefit (Organic Simultaneous LSTM and CNN Fit): A Novel Deep Learning-based Solution for Sentiment Polarity Classification of Reviews. *Expert Systems with Applications*, 2020, 157, 113488. <https://doi.org/10.1016/j.eswa.2020.113488>
13. Kumar, D. A., Chinnalagu, A. Sentiment and Emotion in Social Media COVID-19 Conversations: SAB-LSTM Approach. *Proceedings of the 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, December, 2020, 463-467. <https://doi.org/10.1109/SMART50582.2020.9337098>
14. Levy, O., Goldberg, Y. Neural Word Embedding as Implicit Matrix Factorization. *Advances in Neural Information Processing Systems*, 2014, 27.
15. Li, T. Restaurant Review Data on Dianping.com, 2018.
16. Li, W., Zhu, L., Shi, Y., Guo, K., Cambria, E. User Reviews: Sentiment Analysis Using Lexicon Integrated Two-channel CNN-LSTM Family Models. *Applied Soft Computing*, 2020, 94, 106435. <https://doi.org/10.1016/j.asoc.2020.106435>
17. Liu, L., Lei, M., Wang, H. Combining Domain-Specific Sentiment Lexicon with Hownet for Chinese Sentiment Analysis. *Journal of Computers*, 2013, 8(4), 195-206. <https://doi.org/10.4304/jcp.8.4.878-883>
18. Lu, Z., Du, P., Nie, J-Y. VGCN-BERT: Augmenting BERT with Graph Embedding for Text Classification. *Proceedings of the European Conference on Information Retrieval*, April, 2020, 369-382. https://doi.org/10.1007/978-3-030-45439-5_25
19. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 2013, 26.
20. Mullen, T., Collier, N. Sentiment Analysis Using Support Vector Machines with Diverse Information Sources. *Proceedings of the Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, July, 2004, 412-418.
21. Munikar, M., Shakyia, S., Shrestha, A. Fine-grained Sentiment Classification Using BERT. *Proceedings of the 2019 Artificial Intelligence for Transforming Business and Society (AITB)*, 2019, 1, 1-5. <https://doi.org/10.1109/AITB48515.2019.8947435>
22. Narayanan, V., Arora, I., Bhatia, A. Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model. *Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning*, 2013, 194-201. https://doi.org/10.1007/978-3-642-41278-3_24
23. Ni, R., Cao, H. Sentiment Analysis Based on GloVe and LSTM-GRU. *Proceedings of the 2020 39th Chinese Control Conference (CCC)*, July, 2020, 7492-7497. <https://doi.org/10.23919/CCC50068.2020.9188578>
24. Ouyang, X., Zhou, P., LI, C. H., Liu, L. Sentiment Analysis Using Convolutional Neural Network. *Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015, 2359-2364. <https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.349>
25. Ramadhan, W., Novianty, S. A., Setianingsih, S. C. Sentiment Analysis Using Multinomial Logistic Regression. *Proceedings of the 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*, 2017, 46-49. <https://doi.org/10.1109/ICCREC.2017.8226700>
26. Rehman, A. U., Malik, A. K., Raza, B., Ali, W. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimedia Tools and Applications*, 2019, 78(18), 26597-26613. <https://doi.org/10.1007/s11042-019-07788-7>
27. Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., Matsubara, E. T. BERT for Stock Market Sentiment Analysis. *Proceedings of the 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, 2019, 1597-1601. <https://doi.org/10.1109/ICTAI.2019.00231>
28. Sun, J. Jieba Chinese Text Segmentation, July 2014. <https://github.com/fxsjy/jiebaaccessed>
29. Sun, Y., Wang, S., Li, Y., Feng, S., Wang, H. Ernie 2.0: A Continual Pre-training Framework for Language Understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(05), 8968-8975. <https://doi.org/10.1609/aaai.v34i05.8428>
30. Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling, 2003, 43(6), 1947-1958. <https://doi.org/10.1021/ci034160g>
31. Tan, S., Cheng, X., Wang, Y., Xu, H. Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. *Proceedings of the European Conference on Information Retrieval*, 2009, 337-349. https://doi.org/10.1007/978-3-642-00958-7_31

32. Turc, I., Chang, M-W., Lee, K., Toutanova, K. Well-read Students Learn Better: On the Importance of Pre-training Compact Models. 2019. <https://doi.org/10.48550/arXiv.1908.08962>
33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. Attention is All You Need. ArXiv, 2017, <https://doi.org/10.48550/arXiv.1706.03762>
34. Wu, P., Li, X., Ling, C., Ding, S., Shen, S. Sentiment Classification Using Attention Mechanism and Bi-directional Long Short-term Memory Network. Applied Soft Computing, 2021, 112, 107792. <https://doi.org/10.1016/j.asoc.2021.107792>
35. Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X. Sentiment Analysis of Comment Texts Based on BiLSTM. IEEE Access, 2019, 7, 51522-51532. <https://doi.org/10.1109/ACCESS.2019.2909919>
36. Yan, X., Jian, F., Sun, B. SAKG-BERT: Enabling Language Representation with Knowledge Graphs for Chinese Sentiment Analysis. IEEE Access, 2021, 9, 101695-101701. <https://doi.org/10.1109/ACCESS.2021.3098180>
37. Yang, K., Xu, H., Gao, K. CM-BERT: Cross-modal BERT for Text-audio Sentiment Analysis. Proceedings of the 28th ACM International Conference on Multimedia, October, 2020, 512-528. <https://doi.org/10.1145/3394171.3413690>
38. Yang, L., Li, Y., Wang, J., Sherratt, R. S. Sentiment Analysis for E-commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. IEEE Access, 2020, 8, 23522-23530. <https://doi.org/10.1109/ACCESS.2020.2969854>



This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (<http://creativecommons.org/licenses/by/4.0/>).