International Journal of Information Technologies and Systems Approach
Volume 16 • Issue 2

Sentiment Classification of Social Network Text Based on AT-BiLSTM Model in a Big Data Environment

Jinjun Liu, School of Computing, Yangjiang Polytechnic, China*

ABSTRACT

To tackle the challenge of ineffective sentiment prediction using current sentiment classification methods, this paper introduces a method social network text sentiment classification. The method leverages a bidirectional short and long-term memory model (AT-BiLSTM), specifically designed for a big data environment. First, a vectorized representation of text is realized by introducing a pre-trained BERT model, and the classification results are dynamically adjusted according to the semantic information of the words. Then, the BiLSTM combined with the attention mechanism performs aspect-level sentiment analysis, and the corresponding model AT-BiLSTM is formulated. Finally, the BERT model randomly selects input tags for information masking and pre-trains the proposed model. The proposed method was evaluated against three alternative methods using an identical dataset. The results show that the novel method achieved the highest accuracy, recall, and F1-score, reaching 93.72%, 93.91%, and 92.38%, respectively. Consequently, the proposed method demonstrates superior performance compared to the other three methods evaluated.

KEYWORDS

Attention Mechanism, BERT, Big Data, BiLSTM, Sentiment Analysis

INTRODUCTION

With the booming development of the Internet and social networking sites, social media such as Twitter, Weibo, and WeChat are gradually changing people's lives. More and more people are sharing their experiences on social media, posting their comments and reviews about a product or service, spreading information, and expressing their opinions or feelings about some issues. For instance, they may discuss hot social issues, comment on national policies and laws, and express joy or sadness about their own experiences. These reviews and opinions can be positive, negative, or neutral (Hammou et al., 2019; Khader et al., 2019; Naik et al; 2021). Using big data analysis techniques such as data mining to analyze data generated by social media users and predict their behavior is of great social importance (Han et al., 2019; Chen et al., 2020). Governments and administrations can analyze

DOI: 10.4018/IJITSA.324808
*Corresponding Author

people's emotional responses for hot issues or policies to understand their ideological trends and elicit the prevailing public opinion. Based on users' reviews, companies can analyze their interests to recommend products they may want to buy or to improve the quality of products, while platforms can recommend content that users may be interested in according to their interests so that like-minded users can interact with each other (Hajiali, 2020; Lu et al., 2020; Jena et al., 2019; Alnashwan et al., 2020).

In addition, in the era of information abundance, there is an unprecedented growth in the demand for online public opinion expression, leading to a large amount of information received and transmitted every day or many tweets posted on the Internet or the spread of misinformation. Some of the information may threaten public safety, which is a great challenge to both Internet-related administrations and public and social stability (Jain et al., 2021; Fahd et al.,2021; Deniz et al., 2021). Hence, analyzing the content posted by users on Weibo and actively guiding public opinion to reduce the impact of misinformation are necessary means to maintain societal stability and security (Liu, 2020; Wang & Shin, 2019). Therefore, it is of great research value and practical significance to analyze the content posted by users on social networks and to accurately classify its sentiment.

Sentiment analysis is classifying the sentiments conveyed in documents or statements posted by users as positive, negative, or neutral. There is an enormous amount of useless data on the Internet and sentiment analysis is needed to analyze the data and extract useful information that expresses specific sentimental content (Wang et al., 2019; Seng & Ang, 2019; Correia et al., 2022). However, when analyzing sentiment in Weibo data, it is important not to focus only on the sentences themselves, but also on the information contained in images, retweets, comments and "likes" on the online platforms. In addition, the user's personality and influence on others is also relevant to the sentiment polarity of tweets. Therefore, when the sentiment polarity of the text is not obvious, it is of great significance to take other information into account, such as users' personalities and images, to obtain a more accurate reading of the content's sentiment polarity (Alqarafi et al., 2019; Lappeman et al., 2021; Shi et al. 2020).

This paper proposes a sentiment classification method of social network text based on an attention mechanism and bidirectional short and long-term memory (AT-BiLSTM) model to address the problem of low accuracy of current text sentiment classification methods and difficulty in effective sentiment prediction. Compared with traditional text sentiment classification methods, the main contributions of the proposed method are as follows.

1. The Bidirectional Encoder Representation from Transformers (BERT) model is used to adjust the meaning of words dynamically according to their semantic information during the training process, effectively solving the problem of word polysemy and improving the accuracy of sentiment classification.
2. An attention mechanism is introduced in a BiLSTM neural network, which uses aspect-level word information to predict the sentiment polarity of text.

The remaining chapters of this paper are as follows. Next, relevant research in this field is introduced. The proposed text emotion classification algorithm based on the ATAE-BiLSTM model is described. Then, the experimental procedures are addressed to verify the performance of the proposed model. Finally, the conclusion of the paper summarizes this study.

RELATED WORK

Many scholars have presented studies on social network text sentiment analysis and have achieved good results. Sridharan et al. (2020) presented a platform that can process big data and to develop sentiment analysis systems with the help of dictionaries, available APIs, or external programs. However, this approach cannot apply to the sentiment analysis of continuous and multi-dimensional information.

Yuan (2022) proposed an improved support vector machine (SVM) algorithm that analyzed travellers' sentiments using linear classification and kernel functions and used a Hadoop distributed file system. However, the method was limited to analyzing the travellers' emotions. Raviya and Vennila (2021) designed and developed a hybrid model pipelining a convolutional neural network and an SVM for sentiment feature extraction and classification. However, this method had low accuracy in capturing syntactic and semantic relationships.

To address the problem of intelligent services in identifying data sources, a classification framework for sentiment prediction on streaming services was used by Kilinc (2019) to design a fake account detection service. The sentiment analysis results were visualized based on real-time reports and dashboard components. However, this method does not have sufficient access to contextual information and the classification accuracy is low. Srivastava et al. (2019) proposed a hybrid approach using naive Bayes and random forest models to mine Twitter datasets and performed sentiment analysis on relevant datasets collected from Twitter using the Twitter API. However, the results of this method were relatively simple, with only positive and negative attitudes.

Zhai et al. (2019) proposed a multi-aspect sentiment attention model combined with sentiment resource attention using neural networks to embed sentences with their aspects of them separately. The method can achieve accurate sentiment classification by considering the influence of aspect relations and the contribution of sentiment resources to sentiment polarity. However, the computation speed of the method is slow, and its convergence speed also needs to be further improved. Wang et al. (2021) proposed a deep-learning sentiment classification model based on weak tagging information. On this basis, the performance of sentiment classification was improved. However, the tagging information model introduced in this method cannot achieve high accuracy easily.

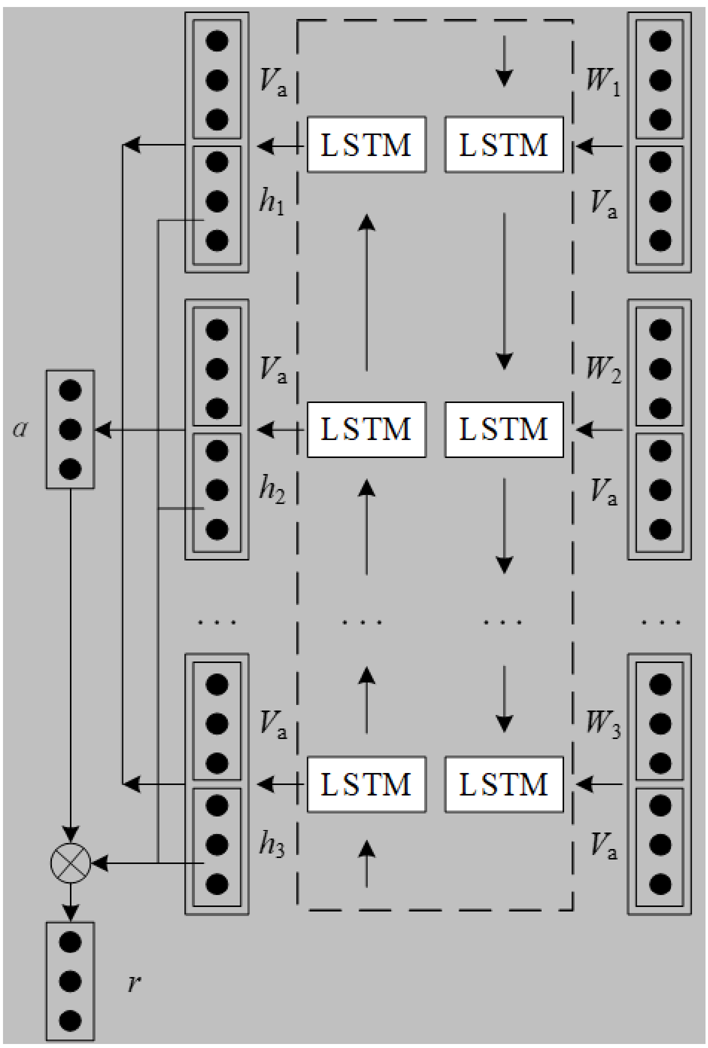
TEXT SENTIMENT CLASSIFICATION ALGORITHM BASED ON AT-BILSTM MODEL

Model Framework

Aspect-level sentiment analysis is a type of deep analysis that determines the sentiment tendency of a particular text aspect. In recent years, unstructured text with personal subjectivity has continued to appear on Internet platforms, and aspect-level sentiment analysis has received great attention from researchers because it allows the accurate analysis of the evaluation objects' sentiment. The sentiment tendency of text is not only determined by its overall semantic information but is also closely related to specific words in given aspects. For example, in a sentence such as "The food is good, but the service is terrible," the polarity of text is positive in the terms of "food" and negative in the aspect of "service." Therefore, it is important to study the connection between specific aspects of words in the sentence and context to analyze the sentiment tendencies of specific targets in the text more accurately. To address the problem that the word embedding vectors' representations are insufficient for sentiment analysis tasks, the text is vectorized using a pre-trained BERT model in the proposed approach. This allows it to dynamically adjust the meanings of words according to their semantic information during the training process and solves the problem of word polysemy. Previous sentiment analysis studies ignored the impact of word aspect information on the whole sentence and did not leverage it to analyze text sentiment tendencies. In this regard, an attention mechanism is introduced and combined with a BiLSTM to perform aspect-level sentiment analysis, which leads to the proposed AT-BiLSTM model. The model combines the aspect word vector and the text vector as the representation for the input vector of the model, and the LSTM network layer extracts the key information of the input vector to formulate the vector representation of the hidden layer. The attention mechanism is introduced to mine the semantic relationships between aspect words and context, and the influence of aspect words on the overall sentiment tendency of the text is also fully considered. During modeling, the attention-based BiLSTM aspect-level sentiment analysis model considers the impact of aspect word information on the whole sentence. The aspect word vector and context vector are combined as the input vector of the model, and then the LSTM network extracts the important information of the input vector from two directions and obtains a vector representation of the hidden layer. The attention mechanism then extracts the semantic relationship between the aspect word vector and the hidden vector to achieve the weighted representation of the text on the specific aspect words. The framework of the model is shown in Figure 1, where V_a denotes the aspect word vector, h_i the hidden layer, W_i denotes the word vector, and α denotes the attention layer. The attention-based BiLSTM aspect-level sentiment analysis model comprises four main components as follows.

1. Model input. The word vectors are obtained using a pre-trained BERT model and are combined with the aspect word vector as the input vector of the model.
2. A bidirectional LSTM network layer. The aspect word vectors and word vectors from the input layer are fed into the LSTM network, and the model extracts the semantic information of words in two directions to obtain the hidden vector representation.

Figure 1. Attention-based BiLSTM aspect-level sentiment analysis model



3. The attention layer combines the hidden state and aspect word vectors. The aspect word vectors are used as model parameters for training to obtain the weighted representation of the text on the given aspect word.
4. Output layer, where the final sentence feature vector is input to a softmax layer to classify emotional polarity.

Pre-Trained BERT Model

The input vector of the AT-BiLSTM model is composed of two parts: the aspect word vector and the text word vector. All word vectors are obtained using a pre-trained BERT model. During training, the BERT model considers the semantic relationship between words and analyzes the semantic information of words in different contexts, thus solving the problem of word polysemy.

The Google open-source project provides two models, which are BERTBase and BERTLarge. The BERTBase model is relatively small in scale, using only a 12-layer Transformer-encoder structure to analyze the semantic information of text, while the BERTLarge model is larger in scale, using a 24-layer Transformer-encoder to extract features from the semantic information of text. In this paper, BERTBase is adopted and used to obtain the word vectors; the relevant parameter settings are shown in Table 1.

The BERT model can be considered as comprising multiple levels, where the input sequence is represented using a combination of the word vector, the word segmentation vector, and the position vector, as shown in Figure 2.

The input encoding process of the model is a sum of word vectors, word segmentation vectors and position vectors, with two special symbols [CLS] and [SEP] added in the process. [CLS] is added in the beginning of the text to indicate the beginning of the sentence, and [SEP] is used to split sentences. The bidirectional network structure of BERT fully learns the internal relations between words, so the representation of the words in the model encapsulates the meaning of other words in the same context and therefore loses the words' predictive significance. To enhance the predictive ability of the network model, BERT randomly selects a portion of the input tokens for information masking and performs prediction analysis on this portion of information prior to training. The final vector representation obtained by the model fuses contextual information from both directions. Similarly, a random selection replaces some information in the input sequence, and BERT understands the semantic relationships between sentences by predicting whether the information at both ends is a continuous text.

BiLSTM Model

LSTM was first proposed in the 1990s. As an improved recurrent neural network, it can significantly enhance dealing with long-term sentence dependencies. There are three main structures set up in the model: the input output gate s_t, the output gate O_t and the forget gate y_t. The basic structure is illustrated in Figure 3.

At moment *t*, the state for each gate can be expressed as

Table 1. Parameter settings for BERTBase

Parameter	Value
Number of network layers	12
Hidden layer size	780
Number of Heads	12
Size of filter	3228

Figure 2. Input sequence representation of BERT model

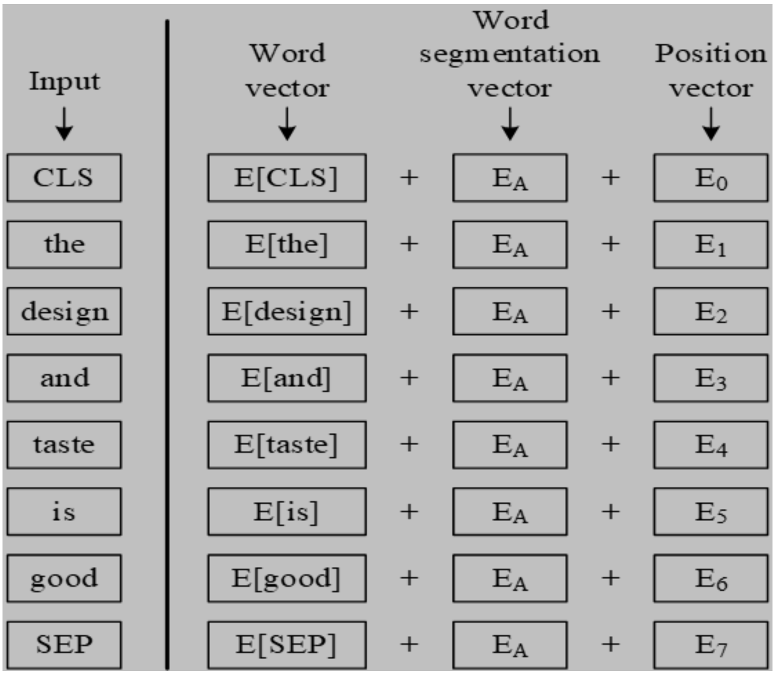
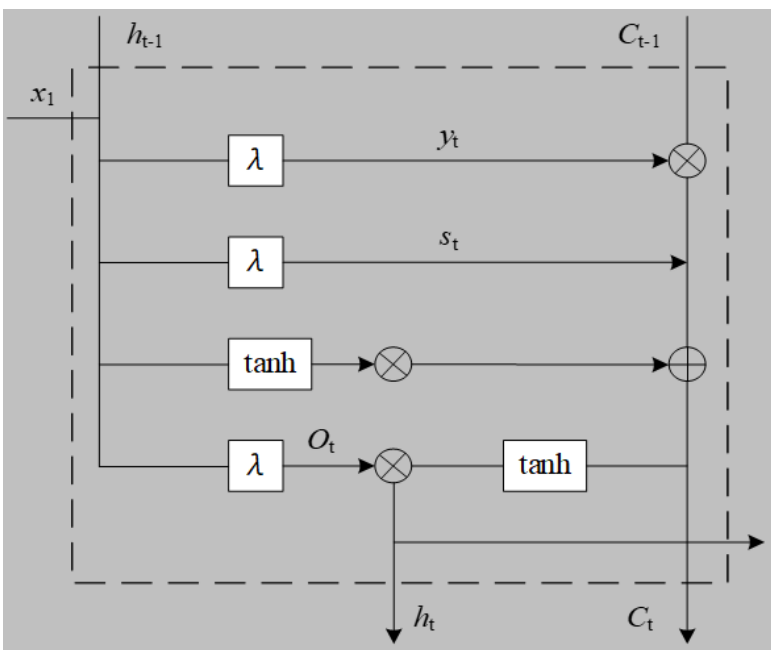


Figure 3. Input sequence representation of BERT model



$$s_t = \lambda\left(\omega_s \cdot \left[h_{t-1}.x_t\right] + p_s\right) \tag{1}$$

$$y_t = \lambda\left(\omega_y \cdot \left[h_{t-1}.x_t\right] + p_y\right) \tag{2}$$

$$O_t = \lambda\left(\omega_O \cdot \left[h_{t-1}.x_t\right] + p_O\right) \tag{3}$$

$$C_t = y_t \cdot C_{t-1} + s_t \cdot \tanh\left(\omega_C \cdot \left[h_{t-1}.x_t\right] + p_C\right) \tag{4}$$

$$h_t = O_t \cdot \tan\left(C_t\right) \tag{5}$$

In Figure 3, s represents the input and h represents the output of the cell. C represents the value of the memory cell, while λ represents the activation function. ω_y, ω_s and ω_O represent the weights of three gates, respectively, while p_y, p_s and p_O represent the corresponding biases. Equations (1), (2) and (3) are used to calculate the input and forget gates respectively. Equation (4) represents the update of the current cell, while (5) is the result of the current output.

The Bi-LSTM is composed of two LSTMs with opposite directions and its structure is shown in Figure 4.

As depicted in Figure 4, the forward and the backward hidden vector of Bi-LSTM at moment *t* can be written as

$$\vec{h}_t = \vec{L}\left(h_{t-1}, \omega_t, C_{t+1}\right), t \in \left[1, T\right] \tag{6}$$

$$\overleftarrow{h}_t = \overleftarrow{L}\left(h_{t-1}, \omega_t, C_{t+1}\right), t \in \left[T, 1\right] \tag{7}$$

$$H_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{8}$$

For example, inputting s_1, s_2 and s_3 to the forward LSTM in turn yields three vectors $\left[h_{01}, h_{02}, h_{03}\right]$, and inputting s_3, s_2 and s_1 to the backward LSTM in turn yields three vectors $\left[h_1, h_2, h_3\right]$. By combining the backward and forward output vectors, $\left\{\left[h_{01}, h_3\right], \left[h_{02}, h_2\right], \left[h_{03}, h_1\right]\right\}$ is obtained. Since the vector $\left[h_{03}, h_3\right]$ contains all the information of the input sequence, it is used as the output vector of the BiLSTM. The whole process is shown in Figure 5.

Attention Model

The attention mechanism can capture key sentence information and assign weighted scores to different words. The hidden state is analyzed with the aspect word vector, which participates in the calculation as part of the model, to obtain a weighted representation of the text on the given aspect word. Let $H \in R^{a \times L}$ be a matrix composed of the hidden vectors $\left[h_1, h_2, h_3, \ldots, h_L\right]$, where a is the size of the hidden layer, L is the length of the sentence, v_n denotes the aspect word vector and $I_L \in R^L$ denotes the vector. An attention weight δ and a weighted hidden vector representation η are calculated as:

$$A = \tanh\left(\begin{bmatrix} A_h H \\ W_n v_n \otimes I_L \end{bmatrix}\right) \tag{9}$$

$$\delta = softmax\left(w^{\mathrm{T}} A\right) \tag{10}$$

$$\eta = A\delta^{\mathrm{T}} \tag{11}$$

where $A \in R^{\left(a+a_d\right) \times L}$. $\delta \in R^L$ and $\eta \in R^a$ are parametric representations. Specifically, δ is a vector consisting of attention weights and η is a weighted representation of the sentence for a given aspect.

Figure 4. Bi-LSTM structure

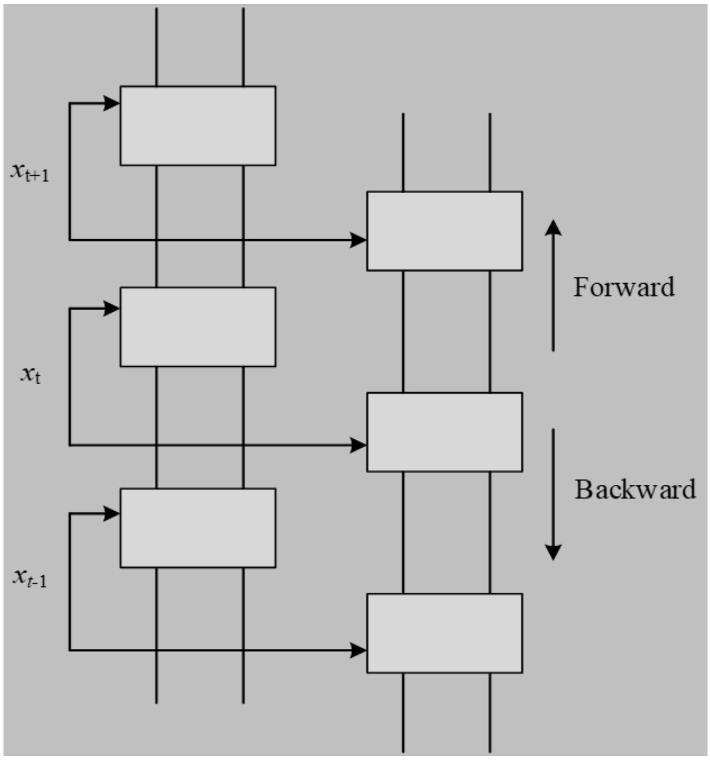
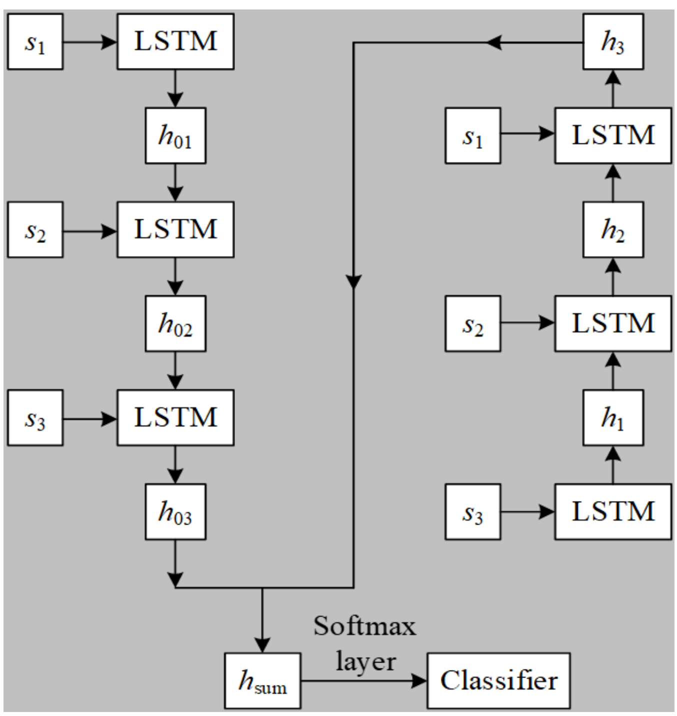


Figure 5. Schematic diagram of the structure of Bi-LSTM



$v_n \otimes I_L$ denotes that the aspect vector v_n has been repeated for L times, where L is the length of the sentence. The final sentence vector can be represented as:

$$h' = \tanh\left(w_1\eta + w_2 h_L\right) \tag{12}$$

$h' \in R^a$, w_1 and w_2 are the parameters to be learned during training. h' represents the sentence feature of the given aspect word. The final sentence vector is input to the softmax layer to obtain the sentiment classification of the aspect word, which can be written as

$$k = softmax\left(W_1 y' + a_1\right) \tag{13}$$

where W_1 and a_1 are the softmax layer's parameters. The model is trained in an end-to-end manner via back propagation, using regularized cross entropy as the loss function. k represents the real classification results of the sentences and y' represents the predicted value of each classification. Model training aims to minimize the error between the real classification results and the predicted value. The loss function can be written as

$$Loss = -\sum_i \sum_j k_i^j \log k'^j_i + \gamma \left\|p\right\|^2 \tag{14}$$

where i denotes the sentence index, j denotes the three classification indices, γ is the L2 regularization coefficient, and p denotes the set of parameters.

EXPERIMENT

Parameter Setting

Simulation experiments were conducted for the proposed sentiment classification method of social network text using the proposed AT-BiLSTM model in a big data environment. The experimental parameters are shown in Table 2.

There are few studies on multimodal sentiment analysis based on Chinese web platforms, and for the sake of protecting user privacy, these platforms have applied measures to avoid data crawling. There are few standard datasets publicly available that can be used for comparison. Therefore, the dataset used for the experiments were images and texts of tweets crawled from Weibo by our team. The process of dataset acquisition was as follows.

1. A Python program was used to crawl the tweets on the platform, including all the information in each tweet (text, images, number of retweets, "likes" and comments) and user information (user profile and history of tweets within ten days).
2. Pre-processing of the tweet data.
3. Tweet data were manually tagged. First, five people tagged the tweets manually, then the emotional polarity of each tweet was determined. If more than two people tagged the same polarity on a tweet, this polarity was assigned to the tweet. In other cases, the emotional polarity of the tweet was considered unclear, and the tweet was removed from the dataset.

The final generated dataset is shown in Table 3.

Table 2. Experimental parameter settings

Parameter	Value
Number of Bi-LSTM layers	3
Number of hidden LSTM layer units	256
Learning rate	0.002
Batch size	64
Iteration rounds	300
Drop rate	0.1
Optimization function	Adam
Loss function	CE Loss

Table 3. Statistical results of Weibo dataset

Emotional Polarity	Amount of Samples
Active	8346
Neutral	6985
Negative	5106
Total	20437

After tagging was completed, the number of followers and the number of comments, "likes" and retweets of each tweet, and the users' authentication information, were extracted to determine the impact factor. The user's historical tweets and personal profiles were extracted to analyze their personality characteristics.

Experimental Results Analysis

Since the pre-trained BiLSTM was used to generate the word vector sequences, the parameters greatly impacted the experimental results. The performance of the AT-BiLSTM model on the dataset under different learning rates to observe the change in the F1-score of the model is illustrated in Figure 6.

As can be seen from Figure 6, the F1-score of the model was greatest when the learning rate was set to 0.002; therefore, all the following experiments were conducted at a learning rate of 0.002.

The ratio of the training, validation, and test sets in the experiment was set to 7:2:1, and the contents of each set were selected randomly. The ADAM optimization algorithm was used during the training process. The performance of convergence speed and accuracy of the proposed model on the dataset as the training step size increased is shown in Figure 7. The value of the loss during training is shown in Figure 8.

As depicted in Figure 7, the accuracy of the proposed model on the training set was relatively low at first. As the number of iterations gradually increased to around 200, the accuracy rose to over 90% and then stabilized, indicating that the model performed well. It can be seen in Figure 8 that the value of loss fluctuated substantially and decreased fast at the beginning of the training period. The loss value gradually stabilized when the number of iterations reached around 200. This stabilization is because the model used the ADAM optimization method to improve the convergence efficiency. The gradual convergence after a certain number of training iterations indicates that the model's performance is stable.

Figure 6. Relationship between F1-score and learning rate

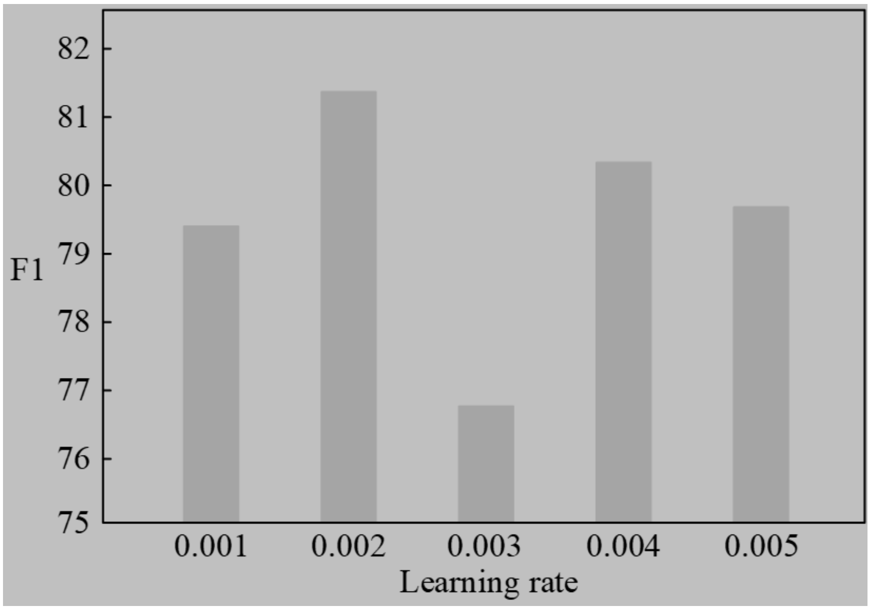
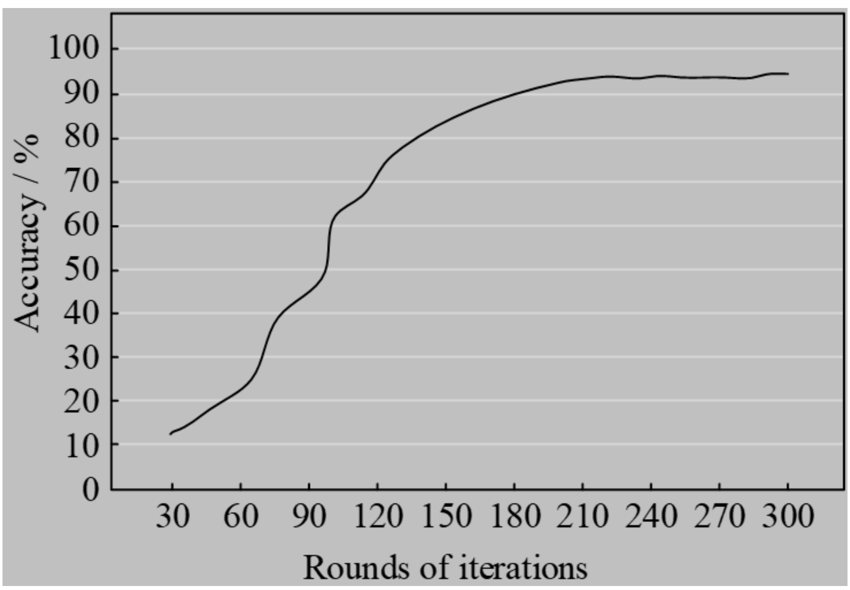


Figure 7. Variation of accuracy during training



Comparative Analysis

In this section, the proposed social network text sentiment classification method based on the AT-BiLSTM model is compared with the methods proposed in Yuan (2022), Raviya and Vennila (2021) and (Zhai et al., 2019). Accuracy, Recall, and F1-score were calculated for the different analysis methods using the same dataset, and the results are demonstrated in Figure 9.

Figure 8. Variation of loss in the training

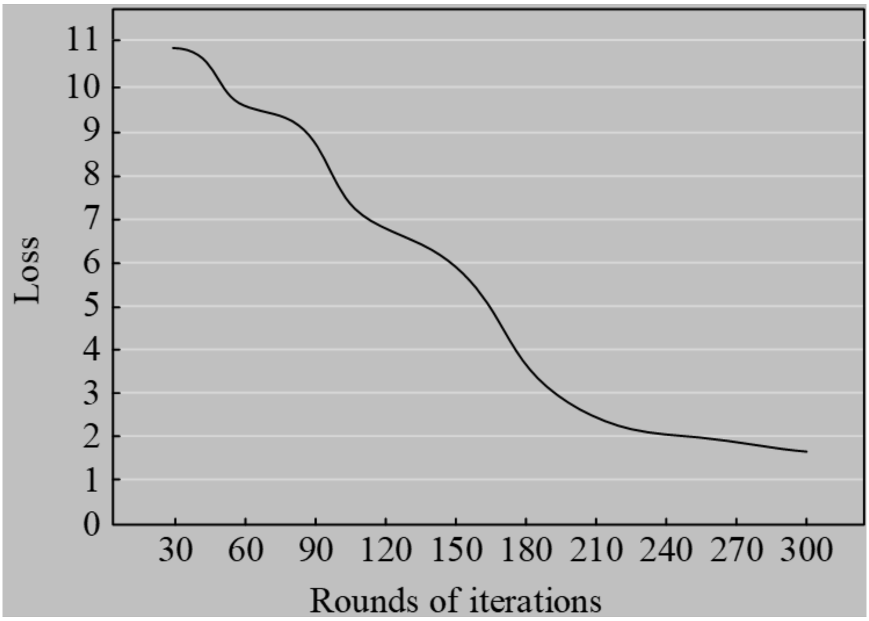
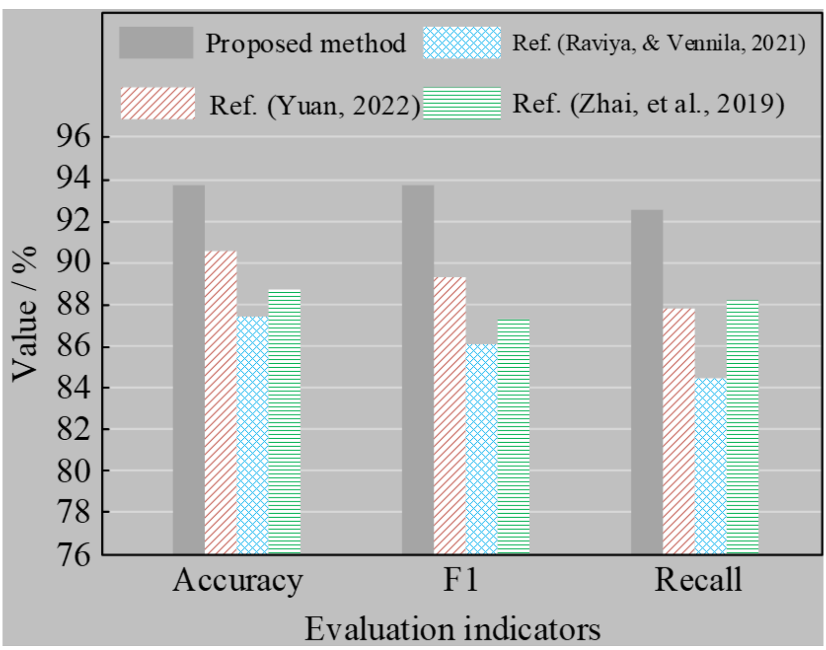


Figure 9. Comparison of different text sentiment analysis methods



As shown in Figure 9, the proposed sentiment classification method outperformed the other three comparison methods in the three evaluation metrics under the same dataset, with Accuracy, Recall, and F1-score reaching 93.72%, 93.91%, and 92.38% respectively. Thus, the proposed method constitutes an improvement over the other three methods. This improvement is because of the introduction of the pre-trained BERT model, which enables dynamic adjustment of the semantic information in the text. Moreover, the AT-BiLSTM model makes full use of information of aspect words to classify the sentiment polarity of the text and combines the aspect word vectors and the text vectors as the representation of the input vector of the model, which achieves a significant improvement in the accuracy of text sentiment classification.

CONCLUSION

To address the problem that current text sentiment classification methods have low classification accuracy and difficulty in effective sentiment prediction, a sentiment classification method of social network text based on the AT-BiLSTM model is proposed. The three basic elements of the proposed method are, first, the word vectors are obtained using a pre-trained BERT to solve the problem of word polysemy. Second, the classification of text sentiment polarity is enhanced via combining of an attention mechanism with the BiLSTM model. Third, the BERTBase is used to obtain the word vectors to improve understanding semantic relationships between sentences. The introduction of the pre-trained BERT model enables dynamic adjustment to the semantic information in the text. Moreover, the AT-BiLSTM model makes full use of information of aspect words to classify the sentiment polarity of the text and combines the aspect word vectors and the text vectors as the representation of the input vectors of the model, which leads to a significant improvement in the accuracy of text sentiment classification.

There are still considerable improvements that can be applied to the proposed method. There are many feature extraction and formulation methods, and different methods may lead to substantial differences in the representation of information or content in some aspects. This paper's feature extraction and calculation methods have not been compared with those of other methods. In future works, such comparisons will be conducted to determine whether performance improvements can be achieved. In addition, future work will focus on rating emotions according to intensity and how to achieve finer-grained emotion analysis to obtain more accurate user emotion classification.

DATA AVAILABILITY

The data used to support the findings of this study are included within the article.

CONFLICTS OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this paper.

FUNDING STATEMENT

This work was not supported by any fund projects.

REFERENCES

Alnashwan, R., Sorensen, H., & O'Riordan, A. (2020). Classification of online medical discourse by modified co-training. *Proceedings of the 5th IEEE International Conference on Big Data Computing Service and Applications (IEEE BigDataService),* 131-137. https://doi.org/ doi:10.1109/BigDataService.2019.00024

Alqarafi, A., Adeel, M., Gogate, A. S., Dashitpour, K., Hussain, A., & Durrani, T. (2019). Toward's Arabic multi-modal sentiment analysis. *Proceedings of the 6th International Conference on Communications, Signal Processing, and Systems (ICCSPS)*, 2378-2386. https://doi.org/ doi:10.1007/978-981-10-6571-2_290

Chen, L. C., Lee, C. M., & Chen, M. Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, *24*(11), 8187–8197. doi:10.1007/s00500-019-04402-8

Correia, F., Madureira, A. M., & Bernardino, J. (2022). Deep neural networks applied to stock market sentiment analysis. *Sensors (Basel)*, *22*(12), 601–609. doi:10.3390/s22124409 PMID:35746192

Deniz, A., Angin, M., & Angin, P. (2021). Evolutionary multiobjective feature selection for sentiment analysis. *IEEE Access : Practical Innovations, Open Solutions*, *9*(32), 142982–142996. doi:10.1109/ACCESS.2021.3118961

Fahd, K., Parvin, S., & de Souza-Daw, A. (2021). A framework for real-time sentiment analysis of big data generated by social media platforms. *Proceedings of the 31st International Telecommunication Networks and Applications Conference (ITNAC)*, 30-33. https://doi.org/ doi:10.1109/ITNAC53136.2021.9652148

Hajiali, M. (2020). Big data and sentiment analysis: A comprehensive and systematic literature review. *Concurrency and Computation*, *32*(14), 205–213. doi:10.1002/cpe.5671

Hammou, B. A., Lahcen, A. A., & Mouline, S. (2019). A distributed ensemble of deep convolutional neural networks with random forest for big data sentiment analysis. *Proceedings of the 5th International Conference on Mobile, Secure and Programmable Networking (MSPN)*, 153-162. https://doi.org/ doi:10.1007/978-3-030-22885-9_14

Han, Z. M., Wu, J. Y., Huang, C. Q., & Zhao, M. (2019). A review on sentiment discovery and analysis of educational big-data. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *10*(1), 146–154. doi:10.1002/widm.1328

Jain, D. K., Boyapati, P., Venkatesh, J., & Prakash, M. (2021). An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification. *Information Processing & Management*, *59*(1), 68–76. doi:10.1016/j.ipm.2021.102758

Jena, R. K. (2019). Sentiment mining in a collaborative learning environment: Capitalising on big data. *Behaviour & Information Technology*, *38*(9), 986–1001. doi:10.1080/0144929X.2019.1625440

Khader, M., Awajan, A., & Al-Naymat, G. (2019). The impact of natural language preprocessing on big data sentiment analysis. *The International Arab Journal of Information Technology*, *16*(3A), 506–513.

Kilinc, D. (2019). A spark-based big data analysis framework for real-time sentiment prediction on streaming data. *Software, Practice & Experience*, *49*(9), 1352–1364. doi:10.1002/spe.2724

Lappeman, J., Clark, J., Evans, R., Sierra-Rubia, L., & Gordon, P. (2021). Studying social media sentiment using human validated analysis. *MethodsX*, *7*(13), 82–90. doi:10.1016/j.mex.2020.100867 PMID:32300546

Liu, B. (2020). Text sentiment analysis based on CBOW model and deep learning in big data environment. *Journal of Ambient Intelligence and Humanized Computing*, *11*(2), 451–458. doi:10.1007/s12652-018-1095-6

Lu, G. Q., Zhao, X. S., Yin, J., Yang, W., & Li, B. (2020). Multi-task learning using variational auto -encoder for sentiment classification. *Pattern Recognition Letters*, *132*(23), 115–122. doi:10.1016/j.patrec.2018.06.027

Naik, M. V., Vasumathi, D., & Kumar, A. P. S. (2021). An adaptable scheme to enhance the sentiment classification of Telugu language. *Social Network Analysis and Mining*, *11*(1), 35–42. doi:10.1007/s13278-021-00764-w

Raviya, K., & Vennila, S. M. (2021). An implementation of hybrid enhanced sentiment analysis system using Spark ML Pipeline: A big data analytics framework. *International Journal of Advanced Computer Science and Applications*, *12*(5), 323–329. doi:10.14569/IJACSA.2021.0120540

Seng, J. K. P., & Ang, K. L. M. (2019). Multimodal emotion and sentiment modeling from unstructured big data: Challenges, architecture, & techniques. *IEEE Access : Practical Innovations, Open Solutions*, *7*(12), 90982–90998. doi:10.1109/ACCESS.2019.2926751

Shi, Y., Zheng, Y. C., Guo, K., & Ren, X. (2020). Stock movement prediction with sentiment analysis based on deep learning networks. *Concurrency and Computation*, *33*(6), 18–25. doi:10.1002/cpe.6076

Sridharan, K., Komarasamy, G., & Raja, S. D. M. (2020). Hadoop framework for efficient sentiment classification using trees. *IET Networks*, *9*(5), 223–228. doi:10.1049/iet-net.2019.0208

Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, *14*(2), 1–16. doi:10.4018/IJHISI.2019040101

Wang, C. T., Yang, X. X., & Ding, L. K. (2021). Deep learning sentiment classification based on weak tagging information. *IEEE Access : Practical Innovations, Open Solutions*, *9*(5), 66509–66518. doi:10.1109/ACCESS.2021.3077059

Wang, G. X., & Shin, S. Y. (2019). An improved text classification method for sentiment classification. *Journal of Information and Communication Convergence Engineering*, *17*(1), 41–48.

Wang, Y., Wang, X. A., Zhang, J. D., & Yu, C. (2019). Combining vector space features and convolution neural network for text sentiment analysis. *Proceedings of the 12th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, 780-790. https://doi.org/ doi:10.1007/978-3-319-93659-8_71

Yuan, Z. (2022). Big data recommendation research based on travel consumer sentiment analysis. *Frontiers in Psychology*, *13*(4), 315–322. doi:10.3389/fpsyg.2022.857292 PMID:35295387

Zhai, G. L., Yang, Y., & Wang, H. (2019). Multi-aspect sentiment attention modeling for sentiment classification of educational big data. *Pattern Recognition and Artificial Intelligence*, *32*(9), 828–834.