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

A GCN- and Deep Biaffine Attention-Based Classification Model for Course Review Sentiment

Jiajia Jiao, College of Information Engineering, Shanghai Maritime University, China
Bo Chen, College of Information Engineering, Shanghai Maritime University, China*
https://orcid.org/0000-0003-0211-0695

ABSTRACT

In recent years, the increasing use of online surveys for course evaluation in schools has led to an outpouring of evaluation texts. These texts, with their emotional polarity, can give schools the most direct feedback. Emotional analysis on course evaluation, therefore, has great implications. However, the not-so-rigid text grammar and rich text content pose a challenge for sentiment analysis in Chinese course evaluation. To solve this problem, this paper proposes a sentiment classification model BiLSTM-GCN-Att (BGAN). Here, BiLSTM is used to extract the features of the text and output the hidden state vector. Then, the deep biaffine attention mechanism is used to analyze the dependence of the text and generate a dependency matrix. Next, input the hidden state vector to the GCN. Finally, the softmax function is used as the output layer of the model to perform sentiment classification. The model proves effective and experimental results, showing that the BGAN achieved a maximum improvement of 11.02% and 14.47% in precision and F1-score respectively compared with the classical models.

KEYWORDS

Chinese MOOC Reviews, Deep Learning, Dependency Parser, Graph Convolutional Network, Sentiment Analysis

INTRODUCTION

With the advent of the information age and the rapid penetration of information digitization into all areas of life and production, the evaluation of college courses, as feedback from students to teachers' courses, is a judgment of teachers' teaching level, students' learning, and the interaction between teachers and students (Liu, 2022).

As a method widely used in universities to collect feedback on the quality of course instruction, Student Evaluation of Teaching (SET) serves as a key ingredient for self-improvement programs, and the feedback solicited from students determines the university administration's decisions, which may affect faculty hiring and course offering plans, among many other things. SET is often used as a reference for faculty promotion and job hiring in universities.

DOI: 10.4018/IJITSA.323568
*Corresponding Author

The massive amount of teaching evaluation data generated after a teaching task is yet to be collated and analyzed for the sentiment. This collation and analysis give the teaching administration a certain understanding and judgment of the teachers' teaching as well as the students' learning level. However, it is impossible to realize the massive amount of data by human power alone, and to accomplish the task requires relying on computers. The task of sentiment analysis can solve this problem exactly (Qiao et al., 2022).

Aspect-level sentiment analysis (ASC) is a branch of natural language processing that belongs to the fine-grained sentiment classification task in the field of sentiment analysis (Chen et al., 2020; Jain et al., 2022; Bao et al., 2021). For example, "Thank you so much!! Your explanation is simply amazing" would be assigned a positive polarity for the aspect explanation.

Sentiment analysis has seen dramatic progress in the last decade or so, but it is mainly intended for the English corpus. The Chinese corpus, or the sentiment analysis of Chinese teaching evaluation, however, is not adequately researched. The sentiment analysis of the Chinese corpus is more difficult most likely because of the Chinese text features: (1) The terms for teaching evaluation at Chinese universities are more diverse and trendy than in the English contexts; (2) Students' preference for concise and direct evaluation expressions requires more efforts in deciphering the relationship between modifiers and subject words; (3) Unlike English, there are no spaces between words in Chinese texts. Because the text needs to be split into words before formal analysis can take place, the Chinese sentences need to be divided into a sequence of words, which makes the usual English corpus model not directly applicable to Chinese (Jia et al., 2019).

This article aims to provide a sentiment analysis of Chinese teaching evaluation with a low false alarm rate. The contributions of this paper are summarized as follows:

We conduct sentiment analysis on course reviews from the Chinese MOOC platform. Based on the BiLSTM model, GCN model and deep biaffine attention mechanism, we propose a Chinese course evaluation model BLGDA, which is tested on comments obtained from the Chinese MOOC. This model proves effective, and it outperforms the classic models such as TextCNN in Precision and F1-score indicators. It can serve as a guide to students' course selection and give timely feedback to teachers as well. It has strong practicability in the field of sentiment analysis for Chinese course evaluation.

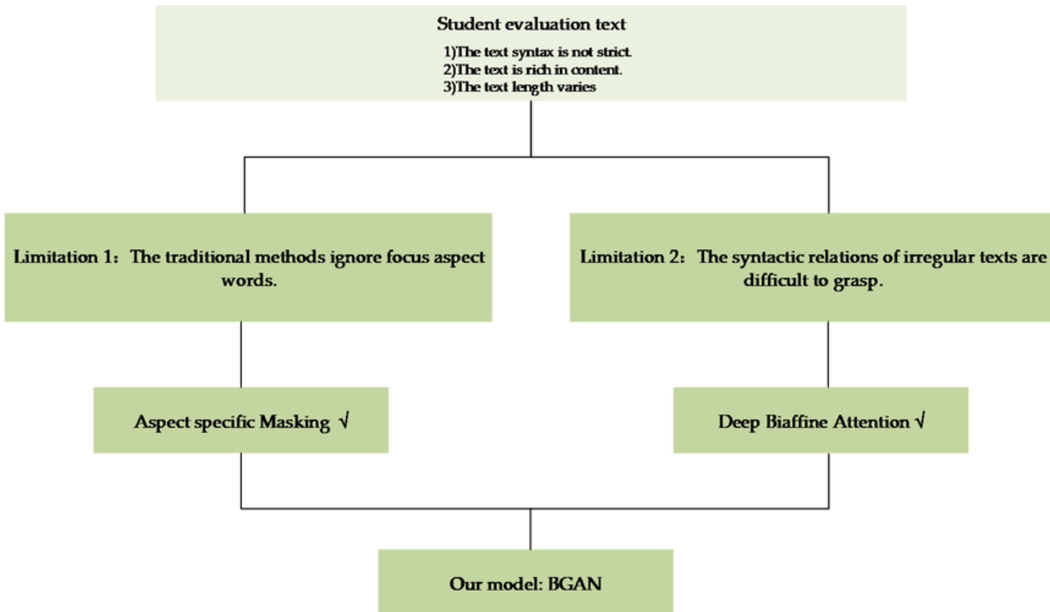
Figure 1 shows the motivation of our work. The biaffine attention mechanism is used in dependency parsing. Compared with traditional dependency parsing methods, biaffine attention can better model the dependencies between words and improve the accuracy of dependency parsing. Biaffine attention can consider global information, not just local information, which helps to improve the accuracy of dependency parsing. Due to the characteristics of the evaluation text, the evaluation text often involves multiple aspects. This model will give an evaluation for a specific aspect, which is helpful for managers to better grasp the emotional tendency of specific aspects, rather than an overall emotional tendency. All in all, the use of deep biaffine attention mechanism for dependency analysis of text and accurately grasp the dependency relationship between text words is helpful to solving the problem of uncritical grammar and unclear referents of course evaluation text. The non-aspect masking layer is used to make the graph convolution operation more focused on the target aspect words and solve the following problems: that the teaching evaluation text is rich in content and the evaluation involves multiple aspects.

The rest of this paper is structured as follows. Section 2 describes the current research related to this study. In Section 3, the GCN and BiLSTM models and the proposed hybrid model BGAN are described. In Section 4, the performance of the proposed algorithm is analyzed and compared with that of several other text classification algorithms. Finally, Section 5 provides a summary.

RELATED WORK

In recent years, great progress has been made in the field of aspect-level sentiment analysis, which can be divided into sentiment analysis methods based on sentiment dictionaries (Du et al., 2022),

Figure 1. The motivation of our work



sentiment analysis methods based on traditional machine learning (Gang et al., 2017), and sentiment analysis methods based on deep learning (Wang et al., 2022).

The sentiment analysis method based on the sentiment lexicon matches the corresponding words in the text according to the sentiment words in the sentiment lexicon, and then the weighted calculation obtains the score and determines the category to which the sentiment belongs. There are AFINN lexicon and SentiWordnet for English sentiment lexicon, and Chinese sentiment lexicon cnsenti published by Dalian Li University, Hownet published by Zhiwang, and NTUSD published by National Taiwan University. This approach is more traditional and nowadays depends on the size of the sentiment lexicon, which is less adapted to the current rapidly updated student language (Sivakumar et al., 2017). Today, there are research efforts to improve the performance of the model by combining sentiment dictionaries with methods such as deep learning (Madani et al., 2020; Yang et al., 2020).

Traditional machine learning-based sentiment analysis methods use statistical machine learning-related algorithms for sentiment determination, mainly using K-Nearest Neighbor (KNN), plain Bayesian, support vector machine (SVM) (Lin et al., 2019), and other related methods.

More deep learning-based sentiment analysis methods now come into use. Deep learning-based sentiment analysis methods use deep network models such as CNN (Liang et al., 2022) and RNN for feature extraction, with techniques such as Long Short-Term Memory (LSTM) and attention mechanism to enhance feature extraction of overall and local information of text (Zhang et al., 2022; Kim et al., 2014). Convolutional neural networks are used on pre-trained word vectors to perform sentence-level text classification, showing improved accuracy compared to traditional machine learning methods (Wu et al., 2022). Song (2019) proposed an attention network to model between context and target entities. Zhang et al. (2019) proposed using graph convolutional neural networks to learn feature representations from syntactic dependencies and fuse other types of features for use on aspect-level sentiment analysis tasks. Liao et al. used RoBERTa for sentiment analysis and their use of RoBERTa is based on deep bidirectional Transformer for sentiment analysis (Liao et al., 2021).

Since text sentiment analysis is often conducted in different contextual areas, and there could be significant deviation in text features, text sentiment analysis is usually carried out in a specific area. Wu et al. (2022) used the capsule network for microblogging text sentiment analysis to improve the efficiency of monitoring public opinion. Fu et al. (2022) studied the text of ancient poetry in the emotional ambiguity and found the short text is a problem for text sentiment analysis. Zhang et al. (2022) constructed a sentiment analysis dataset for such service places as the electric power business hall and used sentiment analysis to help the business hall to improve service quality and user experience in a more targeted manner.

Currently, there is a lack of sentiment analysis for Chinese teaching evaluation and many models intended for English corpus cannot be directly applied to Chinese corpus. In addition, the text in Chinese teaching evaluation is featured by short length, rich expression content, not-so-rigid grammar, and unclear reference. The language habits of the student group give the text data some features distinct from those in other fields. This paper proposes a specific aspect-oriented sentiment classification model BGAN, which uses BiLSTM to obtain text context information from the front and rear directions and outputs hidden vectors, uses the deep biaffine attention mechanism to analyze text dependencies and calculates the scores of different dependency arcs to build a dependency tree, and then inputs hidden vectors to the GCN with a non-aspect word masking layer to obtain aspect features, and finally inputs the aspect features to the softmax function for sentiment classification.

PROPOSED MODEL

This section details the proposed model BGAN for Chinese language teaching evaluation. Figure 2 shows an overview of BGAN model which mainly consists of BiLSTM network, dependent parser, GCN with a non-aspect word masking layer, and sentiment classifier. The text is generated by the BiLSTM to generate hidden state vectors, then sent to the dependent parser using deep biaffine attention for intra-sentence dependency analysis in order to facilitate the subsequent GCN model for graph convolution operations. Finally, text representation is given to the sentiment classifier for classification.

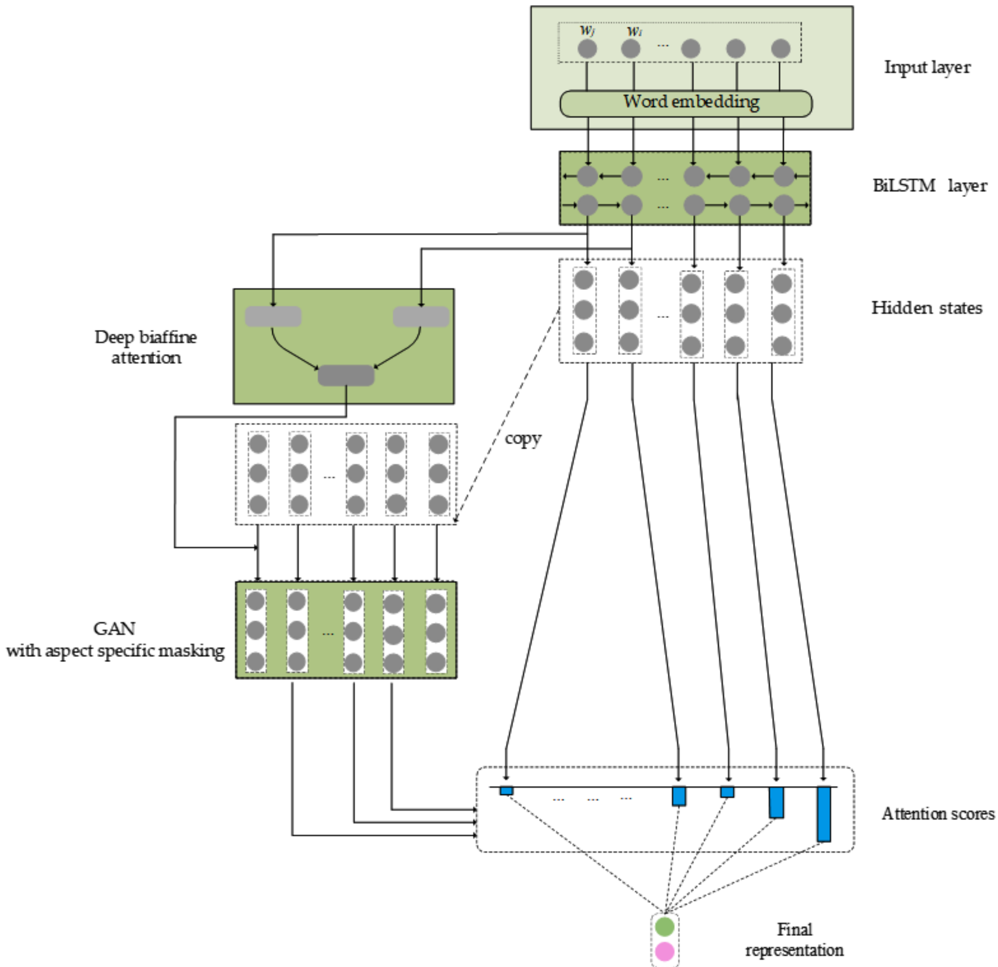
Bidirectional LSTM

Given an n-word sentence $C = (w_1 w_2, \cdots, w_{\tau+1}, \cdots w_{\tau+m}, \cdots, w_{n-1} w_n)$, where there are m-words aspect as $w_{\tau+1}, \cdots w_{\tau+m}$. First embed each word token into a low-dimensional real-valued vector space with embedding matrix $E \in R^{|V| \times d_e}$, where d_e is the number of all words and is the dimension of the word embedding. Where $|V|$ is the number of all words and d_e is the dimension of the word embedding.

In recent years, deep learning networks have been rapidly developed and typical deep network learning models include convolutional neural networks (CNN), recurrent neural networks (RNN), recurrent convolutional neural networks (RCNN), and so on. Student evaluation terms are richly expressed, and often the evaluation involves multiple aspects. Traditional RNNs tend to suffer from gradient dispersion in the presence of long sequence data, which leads RNNs to have only short-term memory; in other words, when RNNs encounter long sequence data, the model can only remember the recent information of the sequence and forget about the earlier input text, thus causing information loss. The LSTM model is then proposed to overcome this problem.

BiLSTM, as a variant of Recurrent Neural Network (RNN), consists of forward LSTM and backward LSTM. Based on the output of both forward LSTM and backward LSTM, BiLSTM integrates the advantages of LSTM to memorize the information of longer distance between texts to obtain deeper text features (Huang et al., 2015), as shown in Figure 3. For the sequence $C = (w_1 w_2, \cdots, w_{\tau+1}, \cdots w_{\tau+m}, \cdots, w_{n-1} w_n)$, the input sequence of forward LSTM of BiLSTM is $(w_1 w_2, \cdots, w_{\tau+1}, \cdots w_{\tau+m}, \cdots, w_{n-1} w_n)$, and the input sequence of backward LSTM is $(w_n w_{n-1}, \cdots, w_{\tau+m}, \cdots w_{\tau+1}, \cdots, w_2 w_1)$, after the network hidden layer extracts the text features to obtain the forward hidden state vector $\{h_{R1}, h_{R2}, h_{R3}, \cdots, h_{Rn}\}$ and the backward

Figure 2. The structure of BGAN framework diagram



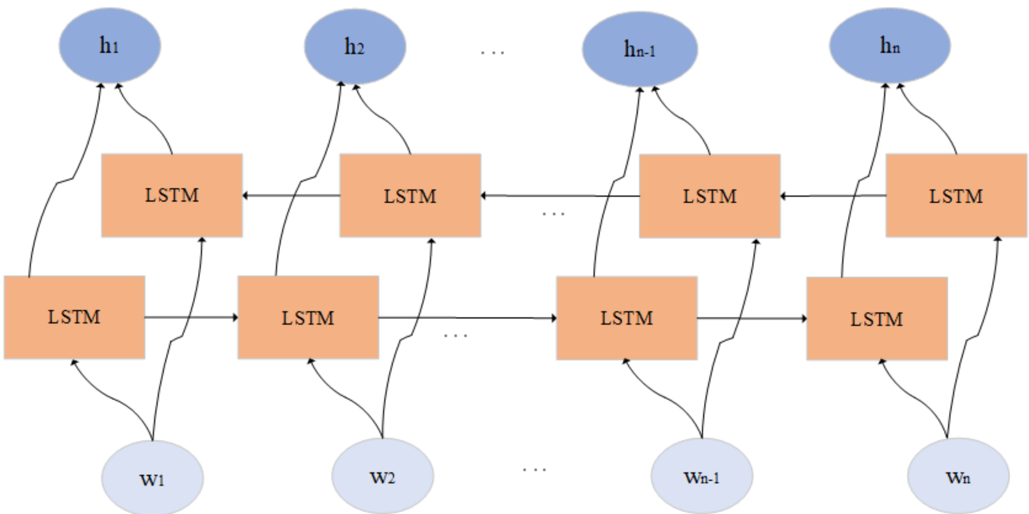
hidden state vector $\{h_{L1}, h_{L2}, h_{L3}, \cdots, h_{Ln}\}$. Then the forward hidden state vector and the backward hidden state vector are stitched to obtain $\{[h_{L1,}, h_{Rn}], [h_{L2,}, h_{Rn-1}], \cdots [h_{Ln,}, h_{R1}]\}$, which is the final output $\{h_1, h_2, h_3, \cdots, h_n\}$.

Dependent Parser With Deep Biaffine Attention

Dependency syntactic analysis is the identification of interdependencies between different words in a sentence (Kübler et al., 2009). According to the theory of dependency syntax, dependency is a master-slave relationship that exists between words. If a word modifies another word, the modifier is said to be dependent, the modified word is called head, and the grammatical relationship between the two is called dependency. The dependency is represented by a directed arc, the direction of which is from the dependent to the head (Mao et al., 2022; Li et al., 2022).

Dependent parser with deep biaffine attention is a graph-based syntactic analysis model that uses biaffine operations to calculate the score of each arc in the text-directed completion graph, finally using a maximum spanning tree algorithm such that each word points to its highest-rated single dominant word (Dozat et al., 2016; Zhang et al., 2020).

Figure 3. The structure of BiLSTM



This dependent parser can be divided into two modules: the multi-layer perceptron attention module and the Biaffine scoring module. The biaffine dependent parser model is shown in Figure 4.

Multi-layer perceptual attention module: to compute the arc $i \leftarrow j$ as an example, first use the output of the words w_i and w_j in the BiLSTM module, i.e. h_i^{bilstm} and h_j^{bilstm}, and then use MLP^D and MLP^H to obtain the representation of the word $h_i^{edge-dep}$ as a dependent word $h_i^{edge-dep}$ and the word $h_j^{edge-head}$ as a dominant word $h_j^{edge-head}$, respectively:

$$h_i^{edge-head} = MLP^H(h_i^{bilstm}) \tag{1}$$

$$h_i^{edge-dep} = MLP^D(h_i^{bilstm}) \tag{2}$$

Biaffine scoring module: based on the output of the multi-layer perceptual attention module, the scoring is performed using an affine transformation as shown in Eqs. (3)-(5). Finally, a maximum spanning tree algorithm is used such that each word points to its highest-rated single dominant word, whereby the dependency tree is obtained:

$$Biaffine(x_i, x_2) = x_1^T U x_2 + W(x_1 \otimes x_2) + b \tag{3}$$

$$S_{i,j}^{edge} = Biaffine^{edge}(h_i^{edge-dep}, h_j^{edge-head}) \tag{4}$$

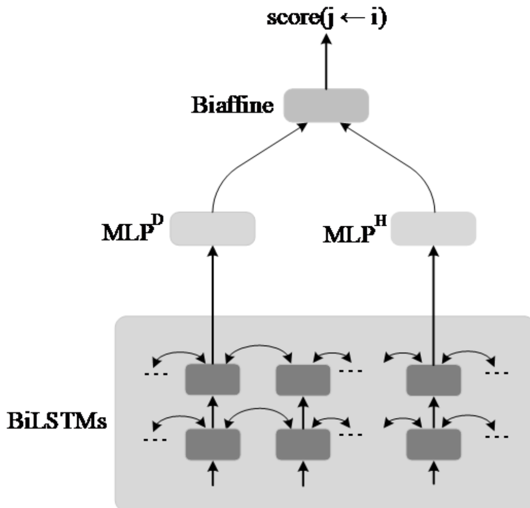
$$score(i \leftarrow j) = sigmoid(S_{i,j}^{edge}) \tag{5}$$

This method is commonly used to construct graphs based on dependency trees, which implies ambiguous information. This is a common method for graph neural networks in NLP. Each token in the text differs in degree, resulting in the representation of the whole sentence being close to the node with a larger degree.

L-Layer GCN With Aspect Specific Masking

Unlike the usual sentiment classification, aspect-oriented sentiment analysis is to make sentiment judgments about an aspect and therefore requires feature extraction around that aspect (Zhang et

Figure 4. Biaffine parser model diagram



al., 2020). In this paper, a multilayer graph convolutional neural network is used on the syntactic dependency tree of sentences and the nodes in each layer are updated with the following formula (Kipf et al., 2016; Zhang et al., 2018):

$$\widetilde{h_t^l} = \sum_{j=1}^{n} A_{tj} w^l g_j^{l-1} \tag{6}$$

$$h_t^l = \mathrm{Re}LU\left(\widetilde{h_t^l} / \left(d_t + 1\right) + b^l\right) \tag{7}$$

where $g_j^{l-1} \in R^{2d_h}$ is the representation of the j-th token in the previous layer of GCN, $h_t^l \in R^{2d_h}$ is the output of node t, $d_t = \sum_{j=1}^{n} A_{tj}$ is the degree of node t in the dependency matrix. W^l and b^t are the trainable parameters.

Afterwards, a position-aware transformation is performed on (Yang. et al., 2017):

$$g_t^l = F\left(h_t^l\right) \tag{8}$$

where $F(\cdot)$ is a weight assignment function (Li et al., 2018), which can give relatively high weights to contexts that are relevant to the aspect words and relatively low weights to contexts that are weakly relevant to the aspect words, which can reduce some noise interference.

The expression of the function $F(\cdot)$ is:

$$q_t = \begin{cases} 1 - \dfrac{\tau + 1 - t}{n} & 1 \le t < \tau + 1 \\ 0 & \tau + 1 \le t \le \tau + m \\ 1 - \dfrac{t - \tau - m}{n} & \tau + m < t \le n \end{cases} \tag{9}$$

$$F\left(h_t^l\right) = q_t h_t^l \tag{10}$$

where q_t is the weight of the t-th token.

The final output of the L-layer GCN is $H^L = \left\{h_1^L, h_2^L, \cdots, h_{\tau+1}^L, \cdots, h_{\tau+m}^L, \cdots, h_{n-1}^L, h_n^L\right\}$, where $h_t^L \in R^{2d_h}$. At this layer the non-aspect hidden state vector is masked out and the state of the aspect word remains constant as shown in Eq. (11).

$$h_t^L = 0 \quad 1 \le t < \tau + 1, \tau + \mu < \tau \le n \tag{11}$$

The output of this zero-masking layer is aspect-oriented features as $H_{mask}^L = \{0, \cdots, h_{\tau+1}^L, \cdots, h_{\tau+m}^L, \cdots, 0\}$, and through graph convolution operations, H_{mask}^L perceives the aspect context in a way that considers both syntactic dependencies as a property and distant inter-word relations. An attention mechanism is used to search for important features related to the semantics of the aspect words, computed as shown in Eqs. (12)-(14), and finally the final predictive representation r, shown in Eq. (15):

$$\beta_t = \sum_{i=1}^{n} h_t^{c\top} h_i^L = \sum_{i=\tau+1}^{\tau+m} h_t^{c\top} h_i^L \tag{12}$$

$$\alpha_t = \frac{\exp(\beta_t)}{\sum_{i=1}^{n} \exp(\beta_i)} \tag{13}$$

$$r = \sum_{t=1}^{n} \alpha_t h_t^c \tag{14}$$

where the dot product is used in Eq. (14) to measure the semantic association between the aspect words and other words for aspect-specific masking, i.e., zero masking.

Emotional Computing Layer

The final representation r of the sentence is obtained and fed to a fully connected layer, and the output of the fully connected layer is normalized by a softmax layer to compute the probability distribution $p \in R^{d_p}$ over the sentiment polarity decision space:

$$p = soft\max(W_p r + b_p) \tag{15}$$

where d_p is consistent with the dimensionality of the sentiment labels, and the parameters W_p, b_p are the weight matrix and bias, respectively. The model is trained using the standard gradient descent algorithm with cross-entropy loss and L2 regularization:

$$loss = -\sum_{(c,\hat{p})} \log p_{\hat{p}} + \lambda \left\|\theta\right\|_2 \tag{16}$$

C is the set of datasets, $\hat{p}$ is the label, $p_{\hat{p}}$ is the probability of label $\hat{p}$, θ represents all trainable parameters, and λ is the regularized hyperparameter.

EXPERIMENT AND ANALYSIS

Dataset

In this paper, we collected over 10,000 datasets from the Chinese MOOC, which are evaluations of different courses, such as computer science and mathematics. These datasets were data cleaned and then data annotated, and the datasets were broken down into three categories, namely positive, neutral, and negative emojis. The pre-processing process is shown in Figure 5.

Positive comments are positive remarks on teachers' teaching attitude, teaching level, PPT production, etc.; neutral comments are neutral remarks on teachers' teaching attitude, teaching level, teaching content, etc. without obvious emotions; negative comments are negative remarks showing their dissatisfaction and disgust for teachers' teaching. Aspect words are related to teaching, such as content, teaching methods, etc. Table 1 shows the statistical information of a dataset.

Some invalid samples and comment samples without obvious aspect words were removed from the data. In this paper, 80% of the datasets were set as the training set and the remaining 20% as the test set. The dataset distribution is shown in Table 2.

In the comment datasets, 42% of the texts are within 10 words and 29% of the texts range from 10 to 20 words, as shown in Figure 6. As can be seen in the table above, the evaluation of the college

Figure 5. Data processing flow

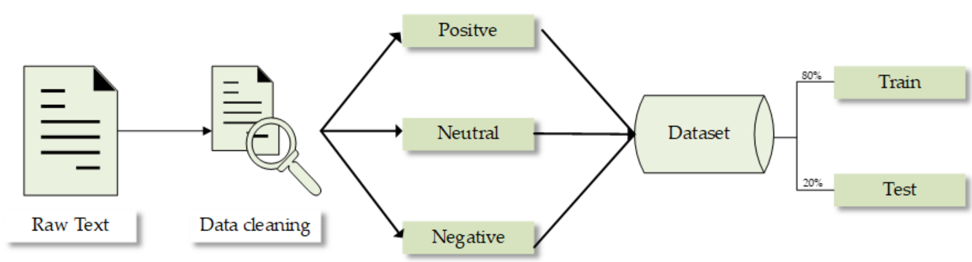


Table 1. Course evaluation dataset sample

Label	Example Evaluation	Aspect Word
Positive	老师讲的内容很好,讲的时候也不拖泥带水,让人收获很大(The teaching content is remarkable, and the teacher is smooth in delivering lectures and I have benefited a lot.)	内容(content)
Neutral	这个课程还行吧(This course is not that bad)	课程(course)
Negative	感觉视频太水了,哪怕每节课多点时间,能讲细致一点,讲多一点,也会觉得没有白听(The video lecture really sucks. If he had invested more time in his lecture, been more specific and prepared more stuff, I would not have felt my time was wasted.)	视频(video)

Table 2. Statistics for the datasets

Label	Training Set	Test Set
Positive	768	192
Neutral	212	53
Negative	8640	2160
Total	9620	2405

student group is mainly based on short texts. It can also be seen that the students' evaluation tends to be direct and explicit, with word cloud statistics conducted on the texts showing that "非常好 (very good)" appeared more than 700 times. Therefore, it is important to find the relationship between the aspect words and adjectives accurately when conducting text analysis.

The deep biaffine attention mechanism in BAN model to parse the dependency relationship between sentences internally focuses more on the relationship between specific aspect words and modifiers, which can effectively improve the accuracy rate.

In summary, the BGAN model achieves better results than other models in sentiment analysis of instructional evaluation.

Experimental Configuration

In this paper, the optimal values of the parameters are selected for the experiments after several comparative experiments. The Adam optimizer (Li et al., 2022) can perform update operations on the parameters based on the oscillations of the historical gradients and the real historical gradients after filtering the oscillations. It has a good interpretation and usually requires either no adjustment or small fine tuning, which is well suited to be applied to scenarios with large-scale data and parameters, among other advantages. L2 regularization to avoid overfitting, batch size, and the number of layers of GCN are the settings that make the model achieve the best results. The adjustable parameters for the experiments are shown in Table 3.

Figure 6. Statistics on the word length of comment texts

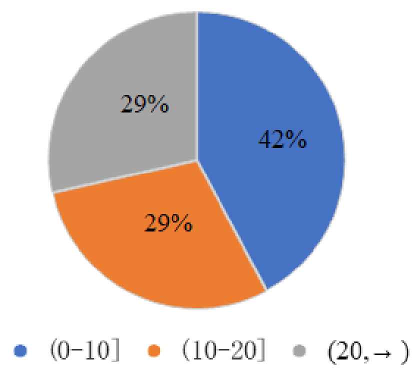


Table 3. Experimental parameter settings

Parameters	Value
Word vector dimension	300
Weight initialization	Uniform Distribution
Hidden state vector dimension	300
Optimizer	Adam with a learning rate of 0.001
Batch size	32
GCN layers	2
L2 regularization factor	$10e^{5}$

Experimental Environment

The experimental platform's hardware configuration and software version are detailed in Table 4.

Metrics

The model evaluation metrics used in this paper are precision, recall, and F1-score. The precision represents the proportion of positive samples that are correctly predicted, the recall represents the stability of the positive proportion of recognition, and F1-score combines the two indicators of precision and recall rate, which can comprehensively reflect the model's classification performance. The better the model's classification performance, the closer the F1-score to 1. The calculation formula is as follows:

(1) TP: the number of comments categorizing positive course reviews as positive.
(2) FP: the number of comments that classify negative course comments as positive.
(3) TN: the number of negative comments classified as negative comments.
(4) FN: the number of comments categorizing positive course reviews as negative.

Precision is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

The F1-score is defined as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{19}$$

Table 4. Experimental platform and pertinent variables

Parameters	Value
OS	Ubuntu
Locales	Python3.6
D-L Framework	PyTorch 1.3.1
Development environment	PyCharm 2022.3.3 (Community Edition)
CPU	Silver 4210 Processor
GPU	[GP106GL] Quadro P2200
RAM	125G
Storage	512GB SSD, 2TB HDD

Comparison and Analysis of Experimental Results

RNN (Wang et al.,2019): With the ability to consider the context, RNN model is suitable for sequence structured data, such as text and audio, which plays a key role in extracting text features.

TextCNN: Yoon Kim proposed to use convolutional neural network to classify text. Firstly, each word is characterized as a vector using the Word2vec model, then it passes through convolution, pooling operations, and finally a fully connected layer followed by a softmax layer is used for classification.

BiLSTM (Graves et al., 2013): The BiLSTM comprises two independent LSTM neural networks, where text sequences are input to two LSTM [38] neural network models in forward and inverse order for feature extraction, so as to facilitate better access to the long-range dependencies between texts.

ASGCN (Zhang et al., 2019): The ASGCN model was proposed by Zhang et al. ASGCN is based on GCN, which can use the features of neighboring nodes to encode and update the node representations in the graph in order to better establish the relationship between words.

As can be seen from the experimental results in Table 5, BGAN outperforms the four baseline models of RNN, TextCNN, BiLSTM, and ASGCN.

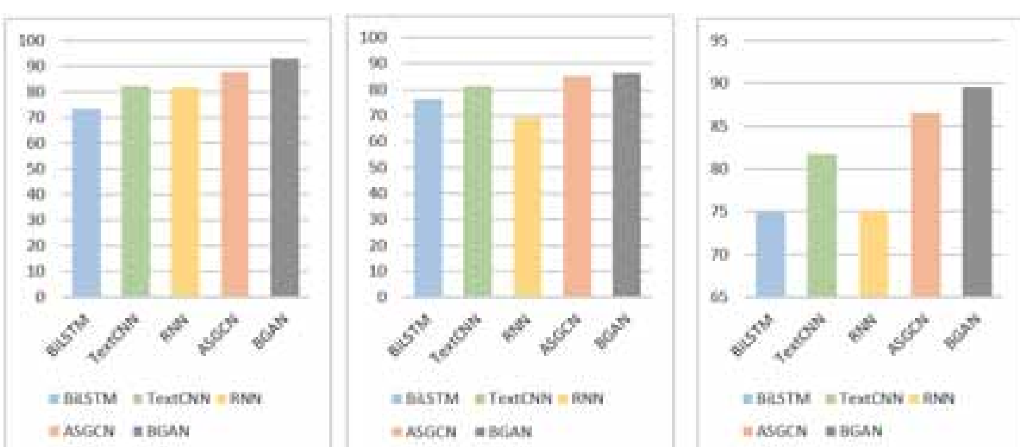
Figure 7-(c) compares the five models in F1 values. Of all the models, the BGAN model achieves the highest F1 value in the comparison experiments, which indicates its improvement in both overall classification performance and the ability to distinguish the three sentiment categories. Thus, the model is proved to have strong sentiment analysis ability.

To verify the effectiveness of the proposed deep learning model for emotional analysis of course evaluation, about 20% of the comments in the dataset are used as the verification set and the remaining data as the training set for experiments. In the experiment on the test set, the change curve

Table 5. Comparison of precision, recall, and F1-score of different models

Model	Precision %	Recall %	F1-Score %
RNN	82.04	69.37	75.17
TextCNN	82.39	81.21	81.80
BiLSTM	73.56	76.37	74.94
ASGCN	87.81	85.38	86.58
BGAN	93.06	86.46	89.64

Figure 7. Comparison of different indicators of experimental results: (a) precision, (b) recall, (c) F1



of precision is shown in Figure 8, and the change curve of LOSS is shown in Figure 9. It can be seen from Figure 9 that LOSS basically remains stable at 1.2 and the model converges, which proves that the model is effective.

Ablation Experiment

The proposed model uses BiLSTM model, GCN model, and deep biaffine attention, because the experimental results show that the BGAN model has better performance than other models. To further explore the influence of each component of BGAN on performance, this paper conducts BGAN ablation experiment, and the realization results are shown in Table 6.

First, when the BiLSTM model (BGAN w/o BiLSTM) is excluded, it can be seen that the model performance has declined in three indicators, indicating that the BiLSTM model extracting text

Figure 8. Precision change curve

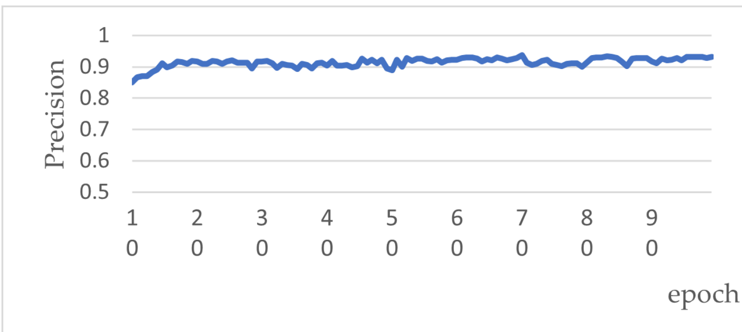


Figure 9. LOSS change curve

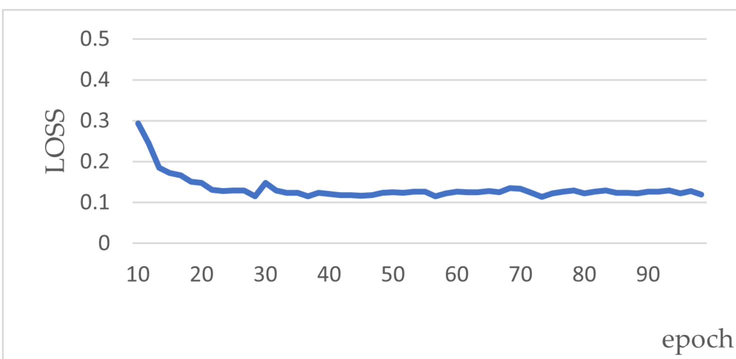
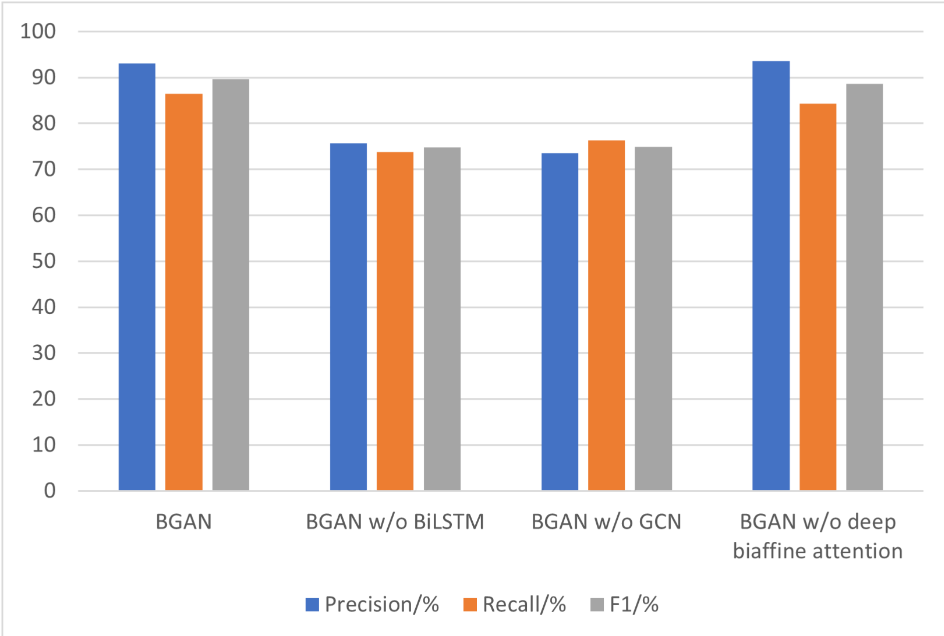


Table 6. Ablation experiment results

Model	Precision %	Recall %	F1 %
BGAN	93.06	86.46	89.64
BGAN w/o BiLSTM	75.73	73.82	74.76
BGAN w/o GCN	73.56	76.37	74.94
BGAN w/o deep biaffine attention	93.56	84.28	88.68

Figure 10. Comparison of ablation experiment results



context features is of great significance to the sentiment analysis of course reviews. Then, with the GCN model removed, the deep biaffine attention mechanism does not need to be used and the entire model is the BiLSTM model. It can be seen that GCN has contributed greatly to the BGAN model, because GCN captures word dependencies and establishes long-term relationships between words. Finally, with the deep biaffine attention mechanism (BGAN w/o deep biaffine attention) removed, it was found that the Precision was slightly better than BGAN, which shows that it is necessary to deeply consider the grammatical relationship in the text to obtain better results for irregular texts such as course evaluation.

Error Analysis

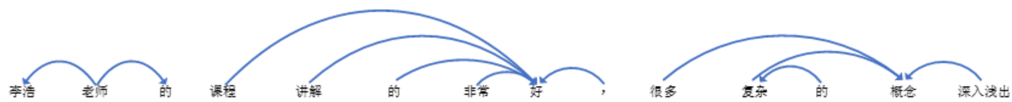
We randomly analyzed 100 misidentified cases. They are roughly divided into three categories: (1) The recognition of neutral words is not accurate enough. The lack of adequate neutral ratings in the training dataset motivates us to collect more neutral ratings to train the model. (2) The text is insensitive to some implicit expressions, such as “let me understand more about the relationship between man and nature.” (3) The judgment of the result is affected by some negative remarks, such as “I have to say that Mr. Wang’s lecture is very good.”

Case Study

The BGAN uses the BiLSTM internally to extract contextual relationship features and the graph convolutional neural network to extract specific aspects of the features, which indicates that the graph convolutional neural network can better handle the relationship features between sentences.

To better understand the text dependency relationship and the classification effect of the model in this paper, we take an example from the test set for visual analysis. As shown in Figure 11, the directed arcs point from the dependent to the head words, representing the relationship between words. In the example text “李浩老师的课程讲解的非常好,很多复杂的概念深入浅

Figure 11. Visualization of dependency analysis of example text



出。(Teacher Li Hao's course explanations are very good, and many complex concepts are explained in simple terms.)", the subordinate word of "讲解 (explanation)" is "好 (good)," which can be judged as its polarity is positive. In addition, "good" is modified by "非常 (very)," which deepens the degree of modification, and "概念 (concepts)" has its modifier "复杂 (complex)." As can be seen from Figure 11, the model can achieve correct analysis when the text is complex in structure and point the words to the corresponding opinion words, thus identifying the emotional polarity of the aspect words.

CONCLUSION

Online course learning has become an important learning method and education informatization has been advancing rapidly. However, there are still many drawbacks with both online and offline courses, and some teachers' teaching methods or the overall quality of the courses still leave much to be desired. The data on students' course evaluation is a strong indicator of the quality of the courses. Sentiment analysis of course evaluations has great implications for the course offering and improving teaching quality at universities.

Considering the ungrammatical and ideographically rich nature of student evaluation terms, traditional models often neglect to dig deeper into the textual dependencies. In this paper, a BGAN model for sentiment analysis of Chinese course reviews is constructed. The model uses BiLSTM model, GCN model, and deep biaffine mechanism. Firstly, BiLSTM is used to extract the features of the text, and then deep bi-affine mechanism is used for the dependency analysis. The maximum spanning tree algorithm is used to generate the dependency tree and the dependency matrix is derived accordingly, which is fed to GCN for further feature extraction combined with the dependency matrix. Finally, the sentiment classification is performed by softmax function. The experimental results show that the model has good stability and great accuracy for sentiment discrimination of Chinese course evaluation.

There are still some deficiencies in this model that can be further optimized. For example, some comments involve multiple languages and some evaluation texts contain emoticons. These are the characteristics of student groups. Mining the emotions of these aspects can further improve the Precision of the model.

In future work, we will expand the amount of data for deeper research and start from the common cyberspeak of student groups to enhance the word embedding process for new vocabulary and to mine more sentiment features for analysis.

AUTHOR CONTRIBUTIONS

Conceptualization, J.J. and B.C.; methodology, J.J.; software, B.C.; validation, B.C., and J.J.; formal analysis, J.J.; investigation, J.J.; resources, B.C.; data curation, B.C.; writing—draft preparation, B.C.; writing—review and editing, B.C., and J.J.; visualization, B.C.; supervision, J.J.; project administration, J.J.; funding acquisition, J.J. All authors have read and agreed to the published version of the manuscript.

FUNDING STATEMENT

This research was funded by Shanghai Pujiang Program, grant number 21PJD026.

DATA AVAILABILITY

Not Applicable. The study does not report any data.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bao, T., Ren, N., Luo, R., Wang, B., Shen, G., & Guo, T. (2021). A BERT-based hybrid short text classification model incorporating CNN and attention-based BiGRU. *Journal of Organizational and End User Computing*, *33*(6), 1–21. doi:10.4018/JOEUC.294580

Chen, X., Sun, C., Wang, J., Li, S., Si, L., Zhang, M., & Zhou, G. (2020, July). Aspect sentiment classification with document-level sentiment preference modeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3667-3677). doi:10.18653/v1/2020.acl-main.338

Du, Z., Huang, A. G., Wermers, R., & Wu, W. (2022). Language and domain specificity: A Chinese financial sentiment dictionary. *Review of Finance*, *26*(3), 673–719. doi:10.1093/rof/rfab036

Fu, C. S., Li, P., Wu, Y. D., Zhang, G. Y., & Gao, J. (2022). Research on poetry sentiment analysis based on imagery. *Modern Computer*, *28*(12), 9–16.

Gang, Z. & Zan, X. (2017). Research on the sentiment analysis model of product reviews based on machine learning. *Information Security Research, 3*(2).

Graves, A. (2012). Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37-45.

Graves, A., Jaitly, N., & Mohamed, A. (2013). *Hybrid speech recognition with deep bidirectional LSTM. In 2013 IEEE workshop on automatic speech recognition and understanding*. IEEE.

Jain, A., Bansal, A., & Tomar, S. (2022). Aspect-based sentiment analysis of online reviews for business intelligence. *International Journal of Information Technologies and Systems Approach*, *15*(3), 1–21. doi:10.4018/IJITSA.307029

Jia, Y. F., Li, S., & Wu, R. (2019). Incorporating background checks with sentiment analysis to identify violence risky Chinese microblogs. *Future Internet*, *11*(9), 200. doi:10.3390/fi11090200

Kim, Y. (2014). *Convolutional neural networks for sentence classification*. Eprint Arxiv. doi:10.3115/v1/D14-1181

Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency parsing. *Synthesis Lectures on Human Language Technologies*, *1*(1), 1–127. doi:10.1007/978-3-031-02131-2

Li, M., Lai, G. H., Chang, Y. M., & Feng, Z. Q. (2022). Performance analysis of different optimizers in deep learning algorithms. *Information Technology and Informatization*, (3), 206–209.

Li, S.K., Li, Y., & Li, Z.H., & Zhang, M. (2022). Interdisciplinary dependency syntax analysis based on ternary training. *Journal of Xiamen University*, *61*(4), 638–645.

Li, X., Bing, L., Lam, W., & Shi, B. (2018). *Transformation networks for target-oriented sentiment classification*. 10.18653/v1/P18-1087

Liang, F., Zhang, H., & Fang, Y. (2022). The analysis of global RMB exchange rate forecasting and risk early warning using ARIMA and CNN model. *Journal of Organizational and End User Computing*, *34*(8), 1–25. doi:10.4018/JOEUC.300762

Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. *Applied Intelligence*, *51*(6), 3522–3533. doi:10.1007/s10489-020-01964-1

Lin, Q., Zhu, Y., Zhang, S., Shi, P., Guo, Q., & Niu, Z. (2019). Lexical based automated teaching evaluation via students' short reviews. *Computer Applications in Engineering Education*, *27*(1), 194–205. doi:10.1002/cae.22068

Liu, Z. (2022). Establishing and improving the teaching evaluation system of colleges and universities in the new development stage. *China Higher Education*, (8), 54–55.

Madani, Y., Erritali, M., Bengourram, J., & Sailhan, F. (2020). A hybrid multilingual fuzzy-based approach to the sentiment analysis problem using SentiWordNet. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, *28*(03), 361–390. doi:10.1142/S0218488520500154

Mao, D., Li, H. Y., & Shao, Y. Q. (2022). Semi-supervised cross-domain semantic dependency analysis technology research. *Chinese Journal of Information*, *36*(02), 22–28.

Qiao, F., & Williams, J. (2022). Topic modelling and sentiment analysis of global warming tweets: Evidence from big data analysis. *Journal of Organizational and End User Computing*, *34*(3), 1–18. doi:10.4018/JOEUC.294901

Sivakumar, M., & Reddy, U. S. (2017, November). Aspect based sentiment analysis of students' opinion using machine learning techniques. In *International Conference on Inventive Computing and Informatics (ICICI).* (pp. 726-731) IEEE. doi:10.1109/ICICI.2017.8365231

Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. (2019). *Attentional encoder network for targeted sentiment classification.* 10.1007/978-3-030-30490-4_9

Wang, R., Li, Z., Cao, J., Chen, T., & Wang, L. (2019, July). Convolutional recurrent neural networks for text classification. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-6). IEEE. doi:10.1109/IJCNN.2019.8852406

Wang, Y. J., Zhu, J. Q., Wang, Z. M., Bai, F. B., & Gong, J. (2022). A review of the application of natural language processing in the field of text sentiment analysis. *Computer Applications (Nottingham)*, *42*(04), 1011–1020.

Wu, G. Z., W, F., & Huang, S. C. (2022). Research on sentiment analysis based on word vector and CNN-BIGRU. *Software Guide*, *21*(08), 27–32.

Wu, R. B., Qiao, H., Jia, Y. F., Liu, S. L., Zhang, Z. C., & Liu, Y. (2022). Sentiment analysis of medium and long microblogs based on capsule network. *Signal Processing, 38*(8), 1632 1641. 10.16798/j.issn.1003-0530.2022.08.008

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 23522–23530. doi:10.1109/ACCESS.2020.2969854

Yang, M., Tu, W., Wang, J., Xu, F., & Chen, X. (2017, February). Attention based LSTM for target dependent sentiment classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). Advance online publication. doi:10.1609/aaai.v31i1.11061

Zhang, Bo. (2020). *Research on treebank transformation and application oriented to dependency syntax*. Soochow University., doi:10.27351/d.cnki.gszhu.2020.002605

Zhang, B., Li, C. G., Zhang, D., Lu, B., Zhou, K. B., Zhang, J., Zhu, Q. M., & Chen, X. P. (2022). A sentiment analysis dataset for service places such as power business offices. *Computer Applications (Nottingham)*, *42*(S1), 37–42.

Zhang, C., Li, Q., & Song, D. (2019). *Aspect-based sentiment classification with aspect-specific graph convolutional networks*. 10.18653/v1/D19-1464

Zhang, W., & Wu, Y. (2022). Semantic sentiment analysis based on a combination of CNN and LSTM model. In *International Conference on Machine Learning and Knowledge Engineering (MLKE)* (pp. 177-180). IEEE. doi:10.1109/MLKE55170.2022.00041

Zhang, Y., & Li, T. R. (2020). A review of review-oriented aspect-level sentiment analysis. *Computer Science*, *47*(6), 194–200.

Zhang, Y., Qi, P., & Manning, C. D. (2018). *Graph convolution over pruned dependency trees improves relation extraction*. 10.18653/v1/D18-1244