International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7767-7779

https,//doi.org/10.70517/ijhsa463666

Using the Emotion Dictionary to Construct an Emotion Analysis Model to Analyze the Emotional Color in Modern Chinese Novels

Juan Li[1,*]
[1] Yinchuan University of Energy, Yinchuan, Ningxia, 750100, China
Corresponding authors: (e-mail: 18709511266@163.com).

Abstract Modern Chinese novels contain rich emotional expressions. In this study, a BERT-BiGRU sentiment analysis model incorporating a sentiment lexicon is constructed for parsing the emotional color in modern Chinese novels. The novel text is preprocessed by jieba segmentation technique, the sentiment lexicon is constructed by combining Hownet and SentiWordNet, and the BERT-BiGRU network architecture is integrated to form a sentiment analysis model with bidirectional semantic comprehension capability. The experimental evaluation shows that the BERT-BiGRU model achieves 93.1%, 92.2%, and 92.6% in precision, recall, and F1-score metrics, respectively, which are 7.4%, 20.1%, and 14.3% higher than the GRU model. When applied to novel text analysis, the model successfully draws sentiment change curves, revealing the laws of sentiment flow in different novels. By calculating the Hurst parameter of 1514 modern Chinese novels, it is found that 93.2% of the excellent novels have a Hurst value greater than 0.5, 87% of which are concentrated in the range of 0.52-0.74, which indicates that the novels' emotional dynamics generally have long-range correlation. The study confirms that the BERT-BiGRU model enhanced by the emotional lexicon can effectively capture the emotional veins in novel texts, providing a computational perspective for literary analysis, and revealing the common law of excellent novels in emotional construction, providing a quantitative reference for novel creation and evaluation.

Index Terms Emotion lexicon, BERT-BiGRU model, Modern Chinese novel, Emotion analysis, Long-range correlation, Hurst parameter

I. Introduction

As an important form of literature, modern Chinese novels occupy an important position in the history of Chinese literature [1], [2]. In these novels, the authors show the diversity and complexity of contemporary society through the revelation of social reality and the analysis of human nature, and attract the attention of readers with their unique language style, narrative techniques and choice of subject matter [3]-[6]. Emotional color is an important driving force for modern Chinese novel writers to create, only when the emotion is moved in the middle, can we create a real literary work, emotional color not only allows us to understand the author's intentions and attitudes in a deeper way, but also enhances our feelings and understanding of the article [7]-[10]. However, due to the different literacy levels of readers, some readers have some difficulties in understanding the emotional color in modern Chinese novels, so the use of sentiment dictionary to construct a sentiment analysis model is of great significance in analyzing the emotional color in modern Chinese novels [11]-[14].

Sentiment analysis is a method of parsing and assessing the emotional tendencies in a text through computer technology [15], [16]. Sentiment lexicon plays a key role as an important part of sentiment analysis [17]. An affective lexicon is a resource that contains a series of words and their corresponding affective polarities [18]. It can be used as the basis for sentiment analysis of modern Chinese novels and used to determine the emotional tendency of individual words in the text [19], [20]. By using the sentiment lexicon to construct a sentiment analysis model, we can quantify and understand the expression of emotions in modern Chinese novels, so as to carry out more in-depth sentiment analysis and mining [21]-[23].

The rapid development of sentiment analysis techniques in the field of computational linguistics has provided a new methodological perspective for the quantitative study of literary texts. Literary works, especially novels, as an important carrier of human emotional expression, the flow and change law of emotions contained in them have been the focus of literary researchers. Traditional literary criticism relies on subjective reading experience and qualitative analysis, which is difficult to comprehensively and objectively present the emotional changes of novel texts. The progress of information technology makes automated emotional analysis possible, providing a quantitative tool for exploring the emotional structure of novel texts. In recent years, the sentiment analysis methods based on machine learning and deep learning have shown good application prospects, but they still face many challenges when facing long novel texts with complex structures and diverse emotional expressions. On the one hand, the obscurity, polysemy and context-dependency of emotional expressions in novel texts make it difficult for simple classification models to accurately capture the emotional information embedded in them; on the other hand, the emotional flow of long texts has a specific temporal order law, which requires models that can efficiently model long-distance semantic dependencies. In addition, the unique language expression and emotional vocabulary system in Chinese literature also put forward special requirements for sentiment analysis models. Therefore, how to construct a sentiment analysis model that is suitable for the characteristics of modern Chinese novel texts, and how to explore the intrinsic laws of novel sentiment changes through the model has become a problem to be solved in the current research.

Based on the above background, this study proposes a BERT-BiGRU sentiment analysis model integrating sentiment dictionaries for analyzing the emotional color in modern Chinese novels. By integrating Chinese sentiment dictionary resources with advanced deep learning models, we make full use of the contextual representation capability of BERT and the sequence modeling advantage of BiGRU in order to solve the challenges in the sentiment analysis of novels with long texts. The study first constructs an emotion dictionary and extracts textual emotion features; then utilizes the BERT-BiGRU model for emotion classification and draws novel emotion curves; finally, the long-range correlation of the emotion curves is analyzed by the Hurst parameter to explore the common law of emotion construction in excellent novels. This study not only helps to deepen the understanding of the emotion structure of modern Chinese novels, but also provides new methodological ideas for computational literary criticism and automated text analysis.

II. Construction of BERT-BiGRU Sentiment Analysis Model Based on Sentiment Dictionary

In view of the limitations of the traditional neural network model, and in conjunction with the latest research advances in text sentiment analysis, this paper designs a BERT-Bi GRU model that incorporates sentiment lexicon information for sentiment analysis of modern Chinese novels.

II. A. Text Sentiment Analysis Process

Text Sentiment Analysis, also called Opinion Mining, refers to the analysis and processing of text with emotional color to dig out the emotions contained in the text, and analyze and classify the obtained emotions. As a research direction of natural language processing, text sentiment analysis has the process of opinion analysis, processing, induction and reasoning. The process of text sentiment analysis is shown in Figure 1.

Data preprocessing refers to the deactivation of data and word separation, and common methods include removing invalid characters and data, using word separation tools for word separation processing, and deactivation word filtering. After word separation, the text is extracted from the sentiment words, and the sentiment words are calculated to finally get the sentiment value of the text, and the text is categorized according to the sentiment value.

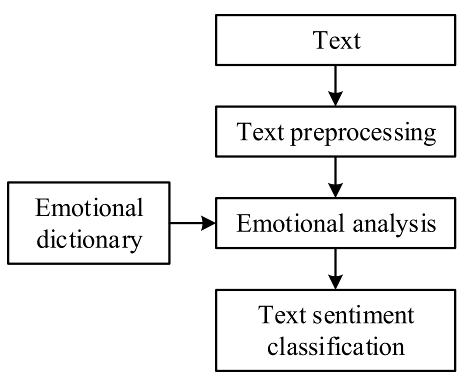

Figure 1: Text sentiment analysis process

This paper analyzes the long texts of modern Chinese novels through the above process of text emotion analysis. Long texts are different from short texts in that they contain complex and varied semantics between emotion words and sentences, so when this paper analyzes long texts, implicit emotion words are added to the emotion dictionary in order to solve the problem of semantic complexity between emotion words and sentences in long texts.

II. B. Steps in the general framework of the model

This paper realizes the sentiment analysis of modern Chinese novels and proposes the corresponding sentiment analysis model framework of BERT-BiGRU [24] that integrates the sentiment lexicon as shown in Fig. 2, which mainly consists of three parts: data processing module, BERT-BiGRU module, and sentiment discrimination module. Its specific realization steps are as follows:

Step 1: Crawl the modern Chinese novel text under the novel reading platform through the open-source Scrapy framework and construct the dataset, and utilize the TIP-LAS lexical system to carry out the data preprocessing steps such as lexical segmentation of the text of the modern Chinese novel and de-duplication of the words and cleaning of the data.

Step 2: Drawing on the rich emotion dictionaries and other resources already available in Chinese, the emotion dictionary is automatically constructed. The steps include collecting emotional vocabulary, screening emotional vocabulary, labeling emotional polarity, and organizing the filtered and labeled emotional polarity vocabulary according to the alphabetical order or other logic to construct the emotional dictionary.

Step 3: The raw data is processed through the data processing layer and embedded into the model through the input layer, while each vocabulary word is mapped to the corresponding vector space. The BiGRU layer further enhances the model's understanding of the contextual semantic features, and utilizes a Softmax classifier to categorize these features. The Softmax classifier is capable of mapping these features to the various emotion categories and outputting the corresponding probability distribution. Ultimately, according to the probability distribution, the sentiment classification result of the text can be determined.

Step 4: Evaluate the performance of the baseline model using accuracy and F1 values.

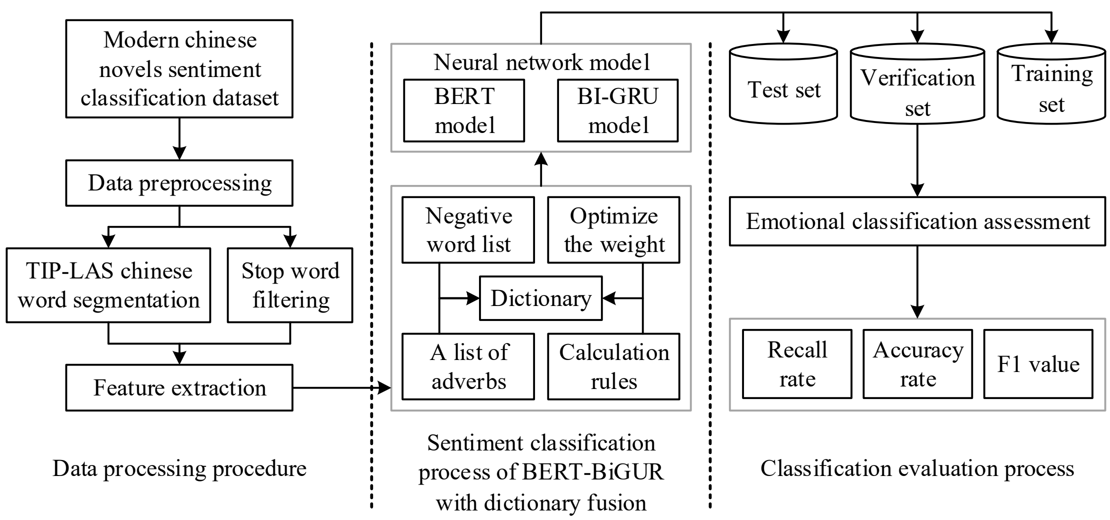

Figure 2: The framework of the BERT-BiGRU model integrating the sentiment dictionary

II. C. Text data preprocessing

The original text obtained through the information acquisition technology can not be directly used for information processing, must be through the text pre-processing text into a convenient computer to identify the structured data, i.e., the text for formal processing. Text pre-processing is the first step in text processing, and its workload accounts for about 80% of the entire text processing process. Therefore, the effect of the text preprocessing process is crucial.

Text preprocessing techniques mainly include word division techniques and removal of deactivated words, processing of outliers, text representation, and feature extraction. Segmentation is the use of word separation tools to divide the text into shorter words. Removing stop words is to remove the pronouns, prepositions, auxiliaries and other function words in the text that generally do not include valid textual properties according to a certain method. Dealing with outliers means that for certain atypical words and texts with fuzzy emotional tendencies that may appear in the modern Chinese novel corpus, the necessary manual verification and labeling processes need to be executed to assist the subsequent emotional classification work.

In this paper, we choose jieba participle technology to analyze the long texts of modern Chinese novels. The python based jieba participle technology is a probabilistic language model participle for Chinese, jieba participle supports three modes: exact mode, full mode and search engine mode, the algorithm is based on the prefix lexicon to realize the efficient word map scanning, and it can generate the with or without loop map to represent all possible word formation situations, and it adopts the dynamic programming algorithm to search for the maximal probability path to find the maximal cut combination based on the frequency of the word. The Jieba lexicon formulates the lexical properties of each participle, and therefore supports returning the lexical properties of the corresponding participle during the word segmentation process. Jieba lexicon also supports keyword extraction to extract keywords from Chinese text that are related to the topic of the text. Keyword extraction is a very important data processing task before doing text clustering, classification, automatic summarization and other data analysis tasks.Jieba thesaurus provides two algorithms, TF-IDF [25] and TextRank [26], to extract keywords from text. Both keyword extraction functions return a list composed of tuples, each tuple contains two elements, the keyword and the corresponding weight.

II. D.Construction of an emotional lexicon

II. D. 1) Dictionary of common emotions

Dictionary-based text sentiment analysis originated from grammar rule-based text analysis, and the following sentiment dictionaries are commonly used in this approach.

(1) Basic Sentiment Dictionary

Hownet Sentiment Dictionary is the commonly used basic sentiment dictionary, GI (The General Inquirer), LIWC (Linguistic Inquiry and Word Count) and so on are also commonly used and mature sentiment dictionaries. When using Hownet, we set the weights for the degree level words.

(2) Expanded Affective Dictionary

Expanding the sentiment lexicon is actually to expand the basic sentiment lexicon by finding the synonyms of the sentiment words through the synonym dictionary.

(3) Domain Dictionary

Dictionary of basic emotion dictionary to identify when in a particular domain some of the emotion words that are not basic also have emotional tendencies. To identify these words the method generally used is the mutual information (PMI) method, if a word and positive words appear together with a high frequency, then the possibility of the word being a positive tendency will also be high, and vice versa. So, by calculating the difference between the frequency of occurrence of a word with a positive word and the frequency of occurrence of a negative word, and setting a certain threshold, one can get a rough idea of the emotional inclination of the word. Calculating co-occurrence can be further subdivided into two methods: one is to calculate the co-occurrence value using search engines, and the other is to calculate the co-occurrence value directly using the corpus.

II. D. 2) The process of constructing an emotion lexicon

In the process of constructing the sentiment dictionary, firstly, the HowNet dictionary can be utilized to obtain the corresponding synonyms of the sentiment words. Secondly, the SentiWordNet database is searched to find and summarize the set of synonyms for each Chinese word. Then, we calculated the average sentiment intensity values of these synonyms to the sentiment tendency intensity values of each synonym. Finally, the average sentiment intensity value of each sense element is calculated to assign a measurable sentiment weight to each sentiment word in the dictionary. In this case, SentiWordnet uses the annotations of the words in the WordNet lexicon, an existing linguistic resource, as the features of the words, classifies the annotated texts, and employs quantitative analyses to determine the positive sentiment, the negative sentiment, and the objectivity weights of the corresponding words in each synonym set.

In the lexicon-based sentiment analysis method in this paper, the sentiment lexicon is extended based on Hownet's sentiment lexicon, and a sentiment lexicon containing 5,396 base sentiment words is obtained through manual screening.

(1) Using Hownet to obtain the corresponding sense elements of words

In HowNet, words often contain multiple sense elements to represent different meanings reflected by words in different contexts. Define a set of word sense elements as W, then, $W=\{M_1,M_2\ldots,M_N\}$, where N denotes the number of sense elements contained in the word, and $M_n(n=1,2,\ldots,N)$ denotes the nth sense element of the word. Let W_Pos be the positive sentiment intensity value of the word and W_Neg be the negative sentiment intensity value of the word, then:

$$W\_Pos=\frac{1}{N}\sum_{n=1}^{N}M\_Pos_n \tag{1}$$

$$W\_Neg = \frac{1}{N}\sum_{n=1}^{N} M\_Neg_n \tag{2}$$

where M_Pos_n is the positive sentiment intensity value of the $S(M)$ th synonym and M_Neg_n is the negative sentiment intensity value of the $S(M)$ th synonym.

(2) Senti WordNet synonym set

In Senti WordNet, a synonym set is the smallest component unit, and a synonym set consists of multiple words with similar meanings. At the same time, each synonym set in Senti WordNet has a corresponding sentiment strength value as the sentiment value labeling, the positive sentiment strength value is denoted by Pos, and the negative sentiment strength value is denoted by Neg. An English word may exist in multiple synonym sets at the same time, indicating that it contains many different semantic concepts. Define the set of synonym sets containing the sense element M as $S(M)$, then $S(M)=\{s_1,s_2,\ldots,s_k\}$, where K denotes the number of synonym sets, and $s_k(k=1,2,\ldots,K)$ is the first one k containing the input sense element. The synonym set of M is searched using the Senti WordNet database for the input semantic elements, and the positive or negative sentiment intensity values of the corresponding synonym set can be obtained.

Each record in Senti WordNet consists of a word number, a lexical property, a positive sentiment value, a negative sentiment value, a synonym entry, and a comment, where the positive sentiment value and the negative sentiment value are between [0,1]. There can be more than one synonym in the synonym entry, each word is represented by "word name #n", and each lexical property and lexical meaning of the same word can correspond to different sentiment values.

(3) Calculate the average sentiment intensity of the set of synonyms

After calculating the average sentiment intensity value of the obtained set of synonyms, the intensity value of the sentiment tendency of each sense element can be obtained. The emotional intensity values of the sense elements can be calculated from equations (3) to (4):

$$M\_Pos = \frac{1}{K}\sum_{K=1}^{K} s\_Pos_k \tag{3}$$

$$M\_Neg = \frac{1}{K}\sum_{K=1}^{K} s\_Neg_k \tag{4}$$

where s_Pos_k is the positive sentiment intensity value of the k th synonym set, and s_Neg_k is the negative sentiment intensity value of the k th synonym set.

(4) Calculate the average sentiment intensity of the sense elements

Calculate the average sentiment intensity value of each sense element to get the sentiment tendency intensity value of the word containing the sense element. Substitute Eq. (3) and Eq. (4) into Eq. (1) and Eq. (2)) to obtain:

$$M\_Pos = \sum_{n=1}^{N}\sum_{K=1}^{K} \frac{s\_Pos_k}{NKn} \tag{5}$$

$$M\_Neg = \sum_{n=1}^{N}\sum_{K=1}^{K} \frac{s\_Neg_k}{NKn} \tag{6}$$

where K_n denotes the number of synonym sets contained in the n th sense element in the input word. After obtaining the above equation, the positive and negative affective tendency values of the corresponding set of synonyms obtained from Senti WordNet can be calculated by substituting them into Eqs. (5) to (6) to find out the positive and negative affective tendency intensity values of the given words. After completing the calculation of all the words, the emotional tendency intensity values of the words are regularized, which can complete the calculation of the emotional intensity values of each word in the process of constructing the emotional dictionary. The extended generated lexicon has a wide range of applications, including predicting the emotional distribution of long texts of modern Chinese novels, recognizing the emotional character of characters, and so on.

II. E. BERT module

BERT [27] is a pre-trained representation model. Its purpose is to obtain a rich representation of textual semantic information from an unlabeled large-scale training corpus, and after several experiments and validations, it is found that the model has no overfitting phenomenon during the training process, thus the drop out mechanism is discarded. Abandoning the drop out mechanism can further improve the model performance and reduce the time complexity of the model.

II. F. BiGRU module

GRU, as a variant of LSTM, has similar capabilities to LSTM in sequence modeling by streamlining and improving the model structure, with fewer parameters than LSTM, but with higher computational efficiency. It is faster on a relatively small corpus. The structure of BiGRU [28] is shown in Fig. 3. Compared to LSTM, the unidirectional GRU has only two gates, and the transfer of states is only a front-to-back unidirectional transfer, which leads to the loss of key information and underutilization of feature vectors.

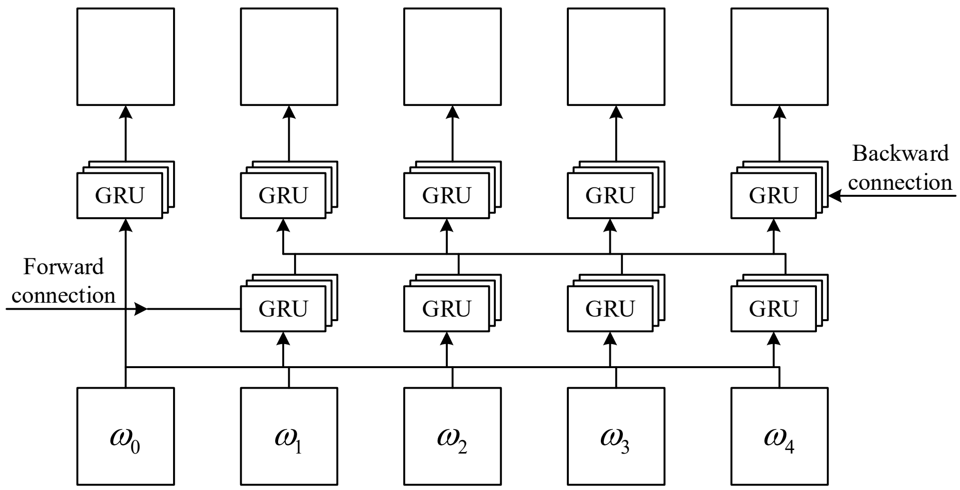

Figure 3: BiGRU structure

In text processing for sentiment categorization, the output result at the current point in time needs to be correlated with the state at the previous point in time as well as the state at the next point in time. This requires the extraction of deep-level feature information of the text, and BiGRU is just able to establish this relationship.

In this case, this paper decides to use the BiGRU network structure.The BiGRU structure allows the network to not only obtain information from the forward propagation, but also to utilize the information in a reverse way, so that more important features can be fully utilized, and thus the features extracted by the network are also richer.Each layer in the BiGRU model consists of two sub-networks, the forward transmission and the backward transmission. Corresponding to the sequential and reverse order of the input sequence respectively. This design is able to capture the forward and backward correlation information in the input sequence. With Eqs. (7) and (8), we can describe these two transmission processes:

$$\overrightarrow{O_c} = \overrightarrow{GRU} = (\{\omega_1, \omega_2, \omega_3 \cdots \omega_{n-1}, \omega_n\}) \tag{7}$$

$$\overleftarrow{O_c} = \overleftarrow{GRU}(\{\omega_1, \omega_2, \omega_3 \cdots \omega_{n-1}, \omega_n\}) \tag{8}$$

Eq. (9) stitches together the forward and backward passes. Such a structure allows the model to simultaneously learn features in the input sequence in different directions, thus improving the expressiveness and performance of the model:

$$O_c = \left[\overrightarrow{O_c}, \overleftarrow{O_c}\right] \tag{9}$$

BiGRU encodes the context by taking the output of the first layer as the input of the second BiGRU layer and so on. Eqs. (10) to (13) represent the hidden layer representation of the context obtained after BiGRU encoding:

$$\overrightarrow{O_c^n} = \overrightarrow{GRU}(\{\omega_1^n, \omega_2^n, \omega_3^n \cdots \omega_{n-1}^n, \omega_n^n\}) \tag{10}$$

$$\overleftarrow{O_c^n} = \overleftarrow{GRU}(\{\omega_1^n, \omega_2^n, \omega_3^n \cdots \omega_{n-1}^n, \omega_n^n\}) \tag{11}$$

$$O_c^n = \left[\overrightarrow{O_c^n}, \overleftarrow{O_c^n}\right] \tag{12}$$

$$O^s = \{O_1, O_2 \ldots O_\varphi, O_{\varphi+1} \ldots O_{\varphi+m} \ldots O_{n-1}, O_n\} \tag{13}$$

where, $\overrightarrow{O_c^n}$ and are $\overleftarrow{O_c^n}$ are the outputs of $\overrightarrow{GRU}$ and $\overleftarrow{GRU}$ of layer n at moment c, $\overrightarrow{O_c}$ is the output result of BiGRU of layer n at moment c, $O_1, O_2 \ldots O_n$ is the hidden layer representation of $\omega_1, \omega_2 \cdots, \omega_n$, respectively, and O^n is the final representation.

II. G. Emotion discrimination module

The classification algorithm used in this paper is the Softmax function, which treats all inputs and outputs as multiple possible classes and can help us turn complex information into a finite probability distribution.

Instead of determining a unique maximum value, Softmax assigns a probability value to each classified output that characterizes the probability that the output belongs to each class. And taking the inputs of i nodes as a starting example, as a way to calculate their degree of similarity, the formula is equation (14):

$$P_y^{(i)} = j \mid x^{(i)}; \theta = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{n=1}^{k} e^{\theta_n^T x^{(i)}}} \tag{14}$$

where θ represents the vector parameters and k represents the number of classification categories. With the Softmax function it is possible to convert the output value of the multiclassification into a probability distribution with probability in the range [0,1] and sum to 1.

The cross-entropy loss function chosen for the loss function, the loss function equation is shown in equation (15):

$$loss = -\sum_{i=1}^{c} k_i \log(\hat{p}_i) + \lambda \| v \|^2 \tag{15}$$

where $\hat{P}$ is the sentiment polarity of the word vector representation, k_i is the actual sentiment polarity, λ is the coefficients of the L2 regularization, and v is the parameter to be trained.

III. Performance evaluation experiments on sentiment analysis models

In order to verify the effectiveness of the proposed BERT-Bi GRU sentiment analysis model in the practical application of sentiment analysis in modern Chinese novels, this chapter conducts performance evaluation experiments on the model.

III. A. Data preparation

In this paper, we select the crawled modern Chinese novel texts to construct the experimental dataset, which has a wide variety of texts and its coverage is extremely wide, and each text is labeled with labels corresponding to the two categories of positive and negative.

III. B. Experimental conditions

The experimental environment configuration is shown in Table 1.

Table 1: Experimental environment

Experimental environment	Configuration
Python	3.12.0
Tensorflow	2.12.0
Kears	2.12.0
Operating system	Centos 8.1
CPU	Intel i7 - 12700
GPU	RTX2080Ti

The specific model parameters are shown in Table 2.

Table 2: Model parameters

Model parameters	Parameter setting
Pre-trained model	BERT-base
Word embedding dimension	128
Hidden layer	768
Batch size	16
Activation function	ReLu
Learning rate	0.0001
Optimizer	Adam
Loss function	Cross-entropy

III. C. Evaluation indicators

In order to evaluate the classification effect of the model, TP is used to denote actual positive samples and predicted positive samples, FP denotes actual negative samples but predicted positive samples, and FN denotes actual positive samples but predicted negative samples, and the precision rate (P), the recall rate (R), and the reconciled mean of the precision rate and the recall rate (F1-score) are used to evaluate the model effect, and the calculation formula is as follows:

$$P=\frac{TP}{TP+FP} \tag{16}$$

$$R=\frac{TP}{TP+FN} \tag{17}$$

$$F1-score=\frac{2PR}{P+R} \tag{18}$$

III. D. Analysis of experimental results

In this experiment, the modern Chinese novel text is divided into training set and test set in the ratio of 8:2, and combined with the five-fold cross-validation, the classification effect of each model is tested on the constructed dataset, and the average value of the results of the five experimental runs is taken, and the specific results of the evaluation indexes of each model are shown in Table 3.

As can be seen from the experimental results, in terms of the index values of precision rate, recall rate and F1-score, compared with the GRU model, the BERT-BiGRU model improves by 7.4%, 20.1% and 14.3%, respectively. And the BERT-BiGRU model improved 3.9%, 2.7% and 3.3% compared to the BERT model and 4.0%, 2.9% and 3.4% compared to the GRU-Attention model. The experimental results show that compared with other models, the BERT-BiGRU model based on sentiment lexicon has a more excellent performance on the sentiment classification of modern Chinese novel texts.

Table 3: Evaluation indicators of each model

Model	P	R	F1-score
BERT-BiGRU	93.1%	92.2%	92.6%
GRU	85.7%	72.1%	78.3%
BERT	89.2%	89.5%	89.3%
GRU-Attention	89.1%	89.3%	89.2%

IV. Emotional coloring empirical analysis in modern Chinese novels

On the basis of confirming the validity of the sentiment analysis model BERT-BiGRU, this chapter selects a modern Chinese novel text and applies the BERT-BiGRU model to conduct an example analysis to explore the emotional color and its changes in the text.

IV. A. Generation of Emotion Curves

In this section, the novel Mansfield Park is taken as an example, and Mr. Sun Zhili's modern Chinese translation version is selected for the study. The BERT-BiGRU model is used to classify the sentiment of the processed novel corpus, and the sentiment curve of the novel is drawn based on the classification results.

The emotional changes in each chapter of the novel are shown in Figure 4. There are 48 chapters in the novel Mansfield Park, and the emotion changes in each chapter are different, with positive values indicating positive emotions and negative values indicating negative emotions, and the larger the absolute value, the more distinct the emotion difference. It can be seen that there are far more positive emotion values than negative emotion values in the novel. Chapter 24 of the novel has the largest emotion difference value, which is more than 150, and this result is closely related to the ups and downs of the plot in the novel. In this chapter, Fanny waits in anticipation for her brother William, whom she has not seen for many years, and her excitement and joy cannot be calmed for a long time. Although they have not seen each other for a long time, their feelings for each other are still warm. They open their hearts to each other and tell each other about their life situations. After years of living at sea, William becomes a good young man who is decent, energetic and cheerful, and is recognized by Sir Thomas. Furthermore, Chapter 46 is the lowest emotional point of the book. In this chapter, both sisters, Maria and Julia, elope with their lovers. Everyone is shocked by the sisters' behavior, and Fanny cannot imagine the emotional trauma and blow to the Bertram family's reputation. After receiving Edmund's letter, she returns to the estate with her sister Susan to care for the devastated Mrs. Bertram. Overall, the chart of emotional changes in the chapters of the novel shows the differences in emotional values in each chapter. However, where the difference in emotional values is relatively small does not actually mean that emotions do not fluctuate or are absent. For example, in Chapter 32, the negative emotion value is as high as 150, while the positive emotion value is even as high as 177, which fully reflects the complex emotional changes of the main character.

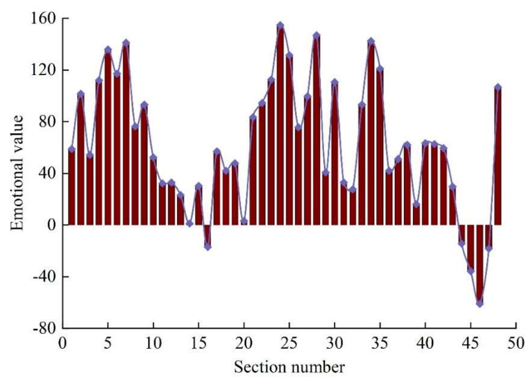

Figure 4: Emotional changes in each chapter

Positive and negative sentiment pairs for each chapter are shown in Figure 5. The highest point of both positive and negative affective values is the last chapter 48, indicating that the affective change of the text reaches its highest point in the last chapter. The positive affective values of all chapters are above 50, and the positive affective values of chapters 7, 25, and 48 are even above 210.Meanwhile, the negative affective values of all chapters except chapters 5, 42, and 43 are greater than or equal to 25, and the negative affective value of chapter 48 reaches above 180. The great changes in positive and negative emotion values reflect the great emotional ups and downs of the novel's characters and emphasize the theme of the novel. Meanwhile, 14 of all chapters have negative affective values less than 50, including chapters 5, 6, 8, 12, 17, 23, 24, 26, 28, 29, 30, 39, 40, 42, and 43. The two chapters with the smallest differences between negative and positive affective values were chapters 14 and 20. For example, Chapter 14 has negative and positive affective values of 71 and 72, respectively, indicating that both opposing emotions are abundant in this chapter.

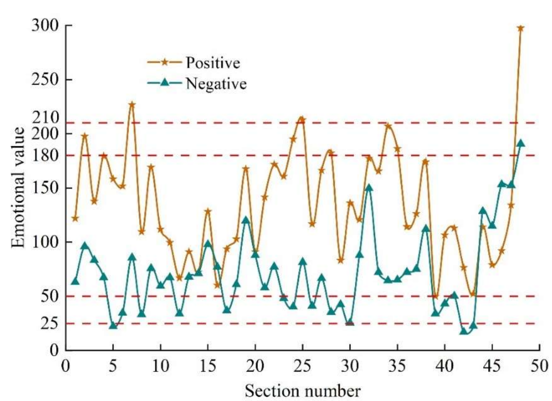

Figure 5: The comparison of positive and negative emotions in each chapter

To summarize, it can be found that the positive and negative emotional values of Mansfield Park are relatively high, and the opposing emotions are rich and have great emotional ups and downs. These changes in emotional values fully reflect the ups and downs of the novel's characters' emotions, and also make the novel's plot more exciting.

IV. B. Emotional color analysis based on the BERT-BiGRU model

The emotion curves of novels are messy and have no obvious change characteristics. In order to further explore the intrinsic value of the emotion curves, this paper interprets the curves with the help of the BERT-BiGRU modeling method, and discovers the common characteristics of excellent novels by exploring the long-range correlation and fractal behavior of the emotion-time series.

Since the initially obtained sentiment curve is non-smooth and contains a lot of noise, which is not conducive to the study, the first step is to remove the impurities in the curve and fit a globally smooth sentiment curve while preserving the variation of the original curve.

The fitted curve plots are still irregular, but they have an overall smoother appearance. Taking Jin Yong's famous novel “The Legend of the Eagle Shooting Heroes” as an example, its emotion curve is plotted as shown in Fig. 6.

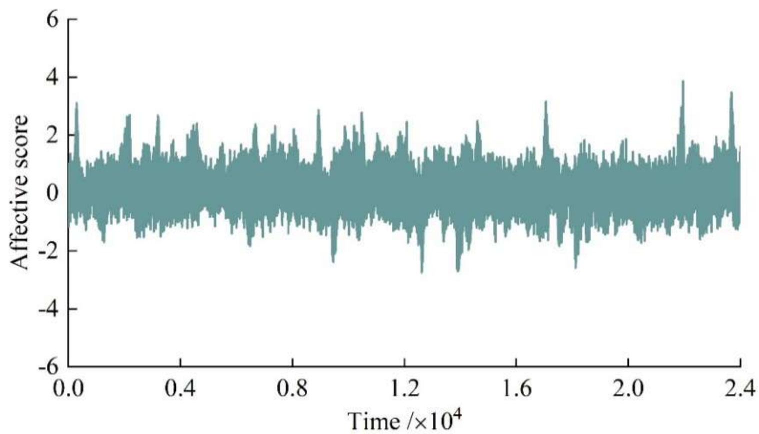

Figure 6: "The Legend of the Condor Heroes" Emotional curve

In order to see the dynamic change of emotion more clearly, the emotion value was adjusted to between [-1,1], and the adjusted emotion curve is shown in Figure 7. Since the orange curve is not smooth enough, it is still impossible to visualize the dynamic change of emotion, so a larger window is chosen to smooth it out, and a cyan curve is obtained, which is the smooth global trend that is sought. Through the black curve, we can see that the initial rendering of the novel is more sad and negative, but there is a tendency to gradually become brighter, with some slight emotional changes occurring in the middle, until finally the whole work is emotionally straightened up, and the ending is framed in a positive scene. This is basically consistent with our reading experience, the protagonist from a nobody gradually rise to fame and the first end of the curve corresponds to the strong rise in emotion at the end corresponds to the Huashan Sword, interspersed with some slight ups and downs in the middle of the plot as a padding and connecting, so that the story development is logical and interesting.

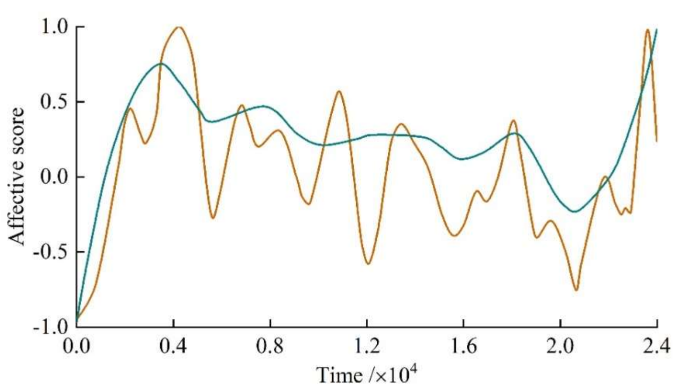

Figure 7: "The Legend of the Condor Heroes" Adjusted emotional curve

The emotion curve of “The Deer Tripods”, which is also a work of Jin Yong, is shown in Figure 8. The adjusted emotional curve is shown in Figure 9. The opening chapter attracts readers with the plot of total turn and then rise, then the development of the story has three twists and turns, although not as good as the opening chapter of the ups and downs of the waves, but also to avoid the article's flat, at a glance, until the final plot returns to calm. The emotional trend of the article coincides with the fate of the main character, and finally Wei Xiaobao returns to the jianghu with his family. The experiment proves that the emotional trend can be more clearly seen through the smooth fitting curve. This also further proves the theory that the emotional changes of a novel can be used as a high proxy for the plot changes.

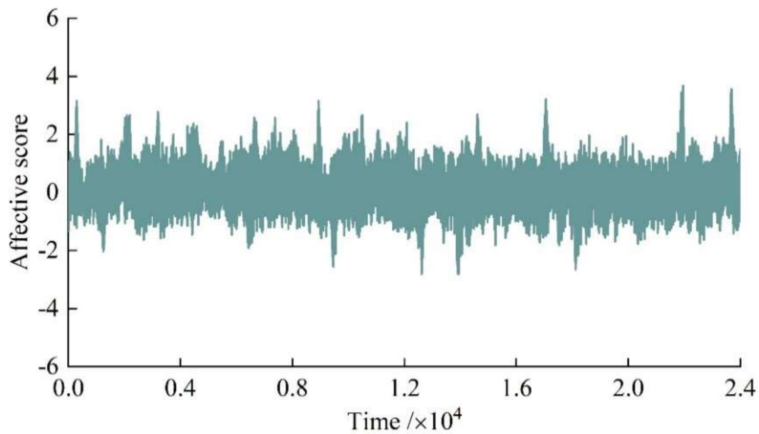

Figure 8: "The Deer and the Cauldron" Emotional curve

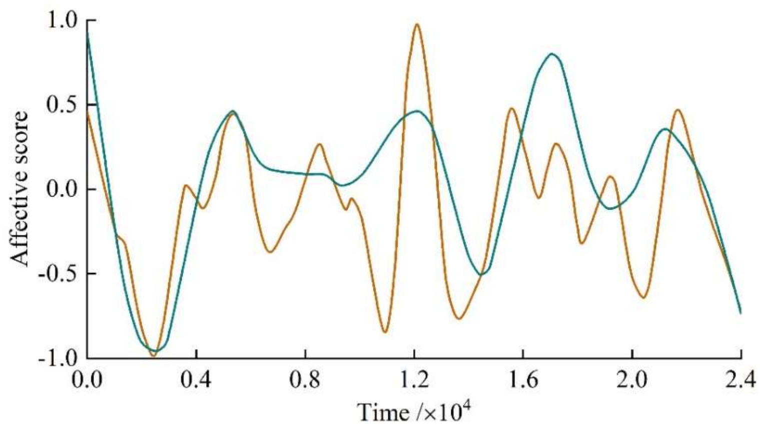

Figure 9: "The Deer and the Cauldron" Adjusted emotional curve

In this paper, 1514 downloaded modern Chinese novels were analyzed as above and the variance (Hurst) values of the residuals under different windows were calculated for each novel, which were divided into three ranges of 0-0.5, 0.5 and 0.5-1 according to the meaning of the Hurst parameter, and the percentages of the different ranges were 3.2%, 3.4%, and 93.2%, respectively. This indicates that 93.2% of the 1514 classic modern Chinese novels studied have Hurst values greater than 0.5, and only 6.6% have Hurst less than or equal to 0.5.

Continued analysis of the novels with Hurst greater than 0.5 shows that among these 93.2% novels, about 87% of the novels are between 0.52-0.74. This indicates that the vast majority of the best novels have sustained long-range correlation in the change of emotional dynamics.

Combined with the meaning of the Hurst parameter, when the Hurst is greater than 0.5, it represents that the trend of emotional change will continue the current direction of development, i.e., the emotional development of the whole article is logical, not abrupt and disorganized, and such emotional change is more easily accepted by readers because it is in line with readers' reading experience and expectations. This provides a mechanism to understand why novels in which the author imagines emotional dynamics can inspire strong emotional resonance in humans.

As a result, this paper finds from the conclusions of the study that there is a commonality of Hurst greater than 0.5, i.e., there is a long-range correlation, and more precisely, the optimal range of this value is between 0.52 and 0.74 for the emotion curves of excellent Chinese fiction works.

V. Conclusion

The BERT-BiGRU model incorporating a sentiment lexicon demonstrates excellent performance in the sentiment analysis of modern Chinese novels. The experiment proves that the F1-score of the model reaches 92.6%, which is 14.3% higher than that of the traditional GRU model, and significantly improves the recall rate while maintaining a high precision rate. The analysis applied to Mansfield Park shows that the model can accurately capture the emotional fluctuations in the novel, e.g., the positive sentiment value of Chapter 24 exceeds 150, reflecting the joy of the protagonist Fanny's reunion with her elder brother; Chapter 46 is the emotional low point of the whole book, and the high value of the negative sentiment reflects the tragic nature of the family crisis in the story. By analyzing the Hurst parameters of 1514 novels, it is found that 93.2% of the excellent novels have a Hurst value greater than 0.5, 87% of which are concentrated in the range of 0.52-0.74, which reveals the inherent law of emotional construction in excellent novels-emotional changes have a continuous long-range correlation. This feature makes the plot development of novels more in line with readers' psychological expectations and enhances emotional resonance.The high efficiency of the BERT-BiGRU model combined with the semantic enhancement of the emotion lexicon provides a computational tool for the study of emotion in modern Chinese novels, and also provides a quantitative basis for the creation and evaluation of novels. Future research can further explore the correlation mechanism between characters' emotions and plot development, and deepen the understanding of novels' emotional structure.

References

[1] Wang, Y. (2015). Fiction in Modern China: Modernity through Storytelling. A Companion to Modern Chinese Literature, 195-213.

[2] Ning, W. (2021). Editor's Introduction: Modern Chinese Literature from Local to Global. Journal of Modern Literature, 44(2), 1-5.

[3] Zhang, Y. (Ed.). (2015). A companion to modern Chinese literature. John Wiley & Sons.

[4] Zhang, C., & Liu, H. (2015). A quantitative investigation of the genre development of modern Chinese novels. Glottometrics, 32, 9-20.

[5] Kong, S. (2016). Diaspora in modern Chinese literature. In The Columbia Companion to Modern Chinese Literature (pp. 62-71). Columbia University Press.

[6] Liu, Y. (2019). Overseas translation of modern Chinese fiction via T'ien Hsia Monthly. Neohelicon, 46(2), 393-409.

[7] Li, J. (2022). Emotion expression in modern literary appreciation: An emotion-based analysis. Frontiers in Psychology, 13, 923482.

[8] Friend, S. (2022). Emotion in fiction: state of the art. British Journal of Aesthetics, 62(2), 257-271.

[9] Khamrabaeva, S. A. (2025). MEANS AND METHODS OF CONVEYING THE EMOTIONAL STATE OF CHARACTERS IN LITERARY TEXTS. SHOKH LIBRARY.

[10] McNamer, S. (2015). The literariness of literature and the history of emotion. PMLA, 130(5), 1433-1442.

[11] Zhao, Q., & Castaneda Abdullah, A. Q. (2024). Metaphorical meanings of color symbols in literature. Chinese Semiotic Studies, 20(4), 625-646.

[12] Lorimer, J. (2023). EACH COLOR REPRESENTS AN EMOTION. A Primer on Arts Integration: Strategies, Lessons, and Collective Wisdom of Teacher Leaders, 79.

[13] Li, J. (2025). Exploring Emotion Analysis in Modern Chinese Fiction Using AI Technology. J. COMBIN. MATH. COMBIN. COMPUT, 127, 1271-1285.

[14] Liu, F., Yang, P., Shu, Y., Yan, F., Zhang, G., & Liu, Y. J. (2023). Emotion dictionary learning with modality attentions for mixed emotion exploration. IEEE Transactions on Affective Computing, 15(3), 1289-1302.

[15] Gangamohan, P., Kadiri, S. R., & Yegnanarayana, B. (2016). Analysis of emotional speech—A review. Toward Robotic Socially Believable Behaving Systems-Volume I: Modeling Emotions, 205-238.

[16] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social network analysis and mining, 11(1), 81.

[17] Hardeniya, T., & Borikar, D. A. (2016). Dictionary based approach to sentiment analysis-a review. International Journal of Advanced Engineering, Management and Science, 2(5), 239438.

[18] Park, Y. J., Kim, S. Y., & Kim, Y. H. (2024). A Child Emotion Analysis System using Text Mining and Method for Constructing a Children's Emotion Dictionary. The Journal of the Korea institute of electronic communication sciences, 19(3), 545-550.

[19] Klinger, R., Kim, E., & Padó, S. (2020). Emotion Analysis for Literary Studies. Reflektierte Algorithmische Textanalyse: Interdisziplinäre (s) Arbeiten in der Creta-Werkstatt. De Gruyter, 237-268.

[20] Sherstinova, T., Moskvina, A., Kirina, M., Karysheva, A., Kolpashchikova, E., Maksimenko, P., ... & Rodionov, R. (2023, May). Sentiment Analysis of Literary Texts vs. Reader's Emotional Responses. In 2023 33rd Conference of Open Innovations Association (FRUCT) (pp. 243-249). IEEE.

[21] Ahmed, M., Chen, Q., & Li, Z. (2020). Constructing domain-dependent sentiment dictionary for sentiment analysis. Neural Computing and Applications, 32, 14719-14732.

[22] Xu, G., Yu, Z., Yao, H., Li, F., Meng, Y., & Wu, X. (2019). Chinese text sentiment analysis based on extended sentiment dictionary. IEEE access, 7, 43749-43762.

[23] Rice, D. R., & Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. Political Science Research and Methods, 9(1), 20-35.

[24] Liu Endong & Lin Junting. (2022). BERT-BiGRU Intelligent Classification of Metro On-Board Equipment Faults Based on Key Layer Fusion. Wireless Communications and Mobile Computing,2022.

[25] Emre Delibaş. (2025). Efficient TF-IDF method for alignment-free DNA sequence similarity analysis. Journal of molecular graphics & modelling,137,109011.

[26] Yongshun Han, Qintu Si & Siriguleng Wang. (2024). Mongolian Automatic Text Summarization Method Based on Pre-trained Model and Improved TextRank. Advances in Computer and Communication,5(2).

[27] Guanzheng Chen. (2025). Enhancing Emotional and Cultural Retention in Ancient Chinese Poetry Translation Using BERT. Asian Journal of Research in Computer Science,18(5),333-343.

[28] Bixiong Luo, Peng Zuo, Lijun Zhu & Wei Hua. (2025). A Wind Power Density Forecasting Model Based on RF-DBO-VMD Feature Selection and BiGRU Optimized by the Attention Mechanism. Atmosphere,16(3),266-266.