International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 3625-3639

https://doi.org/10.70517/ijhsa463306

AI-Driven Ethical Criticism of Literary Narrative: An Example from Chinese New Century Literature

Lihua Dong[1,*]
[1] The School of Culture and Media, Guangdong Cadre College of Science and Technology, Zhuhai, Guangdong, 519090, China
Corresponding authors: (e-mail: weijie0303@163.com).

Abstract The rapid development of Artificial Intelligence (AI) technology has profoundly affected the digitization process in the field of literature, and the application of AI technology in the ethical criticism of digital literary narratives has become more and more widespread. In this paper, combining RoBERTa pre-training model, BiLSTM, and attention mechanism, we constructed a sentiment analysis model based on RoBERTa-BiLSTM-Attention, and analyzed the sentiment trend of the novel text with the object of Chinese New Century Literature "The Riverbank" as a basis for the AI to generate literary narrative ethical criticism. The experimental results show that this paper's model with the multi-head attention mechanism layer improves 3.22% and 1.65% in the two evaluation indexes of accuracy and F1, respectively, compared with the RoBERTa-BiLSTM model without the addition of the multi-head attention mechanism layer, which proves the effectiveness of this paper's model in introducing the attention mechanism. Meanwhile, comparing with other models, the accuracy and F1 value of this paper's model are optimal, reaching 88.35% and 86.52%, respectively, which indicates that this paper's model is suitable for the task of recognizing the emotions of literary texts.

Index Terms Artificial intelligence, RoBERTa, BiLSTM, Attention mechanism, Sentiment analysis, Ethical criticism of literary narratives

I. Introduction

Reflecting on Western literary theories and constructing Chinese contemporary literary theories is not only an academic hotspot, but also a realistic demand [1]. In the past forty years since the reform and opening up, Chinese new century literature, from literary creation to literary criticism, after experiencing a leapfrog development, has already had a worldwide height in many aspects [2], [3]. Since the 1990s, important contemporary Chinese works have been awarded international literary prizes, such as Jia Pingwa's Wasted City in 1997, which won the French Fermina Prize for Foreign Literature, Yu Hua's Knight of the Order of France for Literature and Art in 2004, Mo Yan's Nobel Prize for Literature in 2012, Yan Lianke's Kafka Prize in 2014, Liu Cixin's Three Bodies in 2015, which won the 73rd Hugo Award for Best Long Story, and Cao Wenxuan won the International Hans Christian Andersen Award in 2016 [4]-[7]. The awards of these writers and works have proved the worldwide height of Chinese literature in the new century from the side [8]. Of course, we must soberly realize that the western world's perception of Chinese literature still remains at the level of reading stories and appreciating wonders, and has not really penetrated into the literary and spiritual contents of the works [9], [10]. In other words, the spiritual works provided by Chinese writers for the contemporary world have not been able to exert enough influence on the progress and innovation of the contemporary world's ideology.

Literary criticism also needs creativity. Critics' critical interpretation of contemporary writers and works cannot remain in the application and testing of established theoretical concepts, nor can they be satisfied with the easy-to-understand perceptual appreciation; what is more urgent and important is to complete the production of ideas and theoretical construction through criticism [11], [12]. However, contemporary literary criticism is not satisfactory at the level of thought creation, and this lack is even worse than that of contemporary literary creation [13]. Contemporary literary criticism, which neglects and lacks the creation of ideas, has lagged behind the creation of literature, which binds people's recognition and evaluation of the achievements of contemporary Chinese literature, and also affects the position and reputation of contemporary Chinese literature in the world's literary map [14], [15].

By exploring the future trend of AI-generated literary narrative ethical criticism, this paper identifies a methodological path to promote AI-generated literary narrative ethical criticism through sentiment analysis of digital literary texts. To this end, a sentiment analysis method based on the RoBERTa-BiLSTM-Attention model is proposed. The method obtains a vectorized representation of the text by introducing RoBERTa based on the improvement of

BERT to solve the problem that traditional Word2vec cannot represent the multiple meanings of a word, utilizes the improvement network BiLSTM to make up for the lack of LSTM's inability to utilize the information in the following text, and integrates with the Attention mechanism to highlight the important emotional information in the text. In order to verify the practicality of the model, it is subjected to ablation experiments and compared with other common models, and the ethical criticism of literary narratives is carried out with the example of Chinese new century literature.

II. Future Trends of AI Generative Literary Narrative Ethics Criticism

The iterative updating of big language models and generative artificial intelligence as well as the emergence of human-computer embodied interaction and generation modes have changed the traditional forms of literary text existence, creation and reading, and interpretation concepts.

First, literary text is no longer limited to traditional textual and numerical symbols, it exists in the form of a database, becomes a collection of data that can be retrieved and analyzed, and contains information such as emotional evaluation, meaning analysis, and cultural and social contexts.

Secondly, the function of text has changed, and the digital text supported by database has convertibility, generativity and dialogicity. Conversion is converted from traditional text to digital text or more intuitive visual text form. Generativity highlights the fact that the text in the context of generative AI can dynamically link relevant information, thus promoting the birth of generative AI text, complex hyperlinked text and multimodal text. Dialogicality refers to the fact that in this new text ecosystem, creation and reading are no longer a one-way process, but become a field of interactive communication between human beings and AI intelligences, which strengthens the social interactive value of texts.

Once again, the identities of both writers and critics become plural forms in which many subjects participate. Traditional writers coexist with emerging AI digital human writers who utilize data and algorithms to create, forming a “human-machine-common brain” dual creation mode of textual symbols and digital intelligence, and the addition of AI digital human critics makes literary criticism transcend a single human perspective, presenting a new pattern of diversity and intelligence coexisting. This new pattern not only strengthens the traditional dialogic criticism, but also revolutionizes the generative criticism mode, i.e., the deep interactive dialogue between human beings and intelligences, which stimulates a deeper level of understanding and a circular feedback mechanism.

Finally, in the human-machine embodied interaction scenario based on databases, large language models and generative artificial intelligence, literary activities have evolved into a cooperation, communication and dialogue relationship of "co-creation", "co-reading" and "co-interpretation", which can carry out all-disciplinary, professional, visual dynamics and rich communication and interaction, form an open, trusting and cooperative community of literary activities, and collaboratively promote the innovation and exchange of literary knowledge.

As a result, the structural relationship of literary activities is changing from the "four-element theory" of traditional literary research to the "five-element theory" of literary digital humanities criticism. The Five Elements of Digital Humanistic Criticism in Literature is shown in Figure 1.

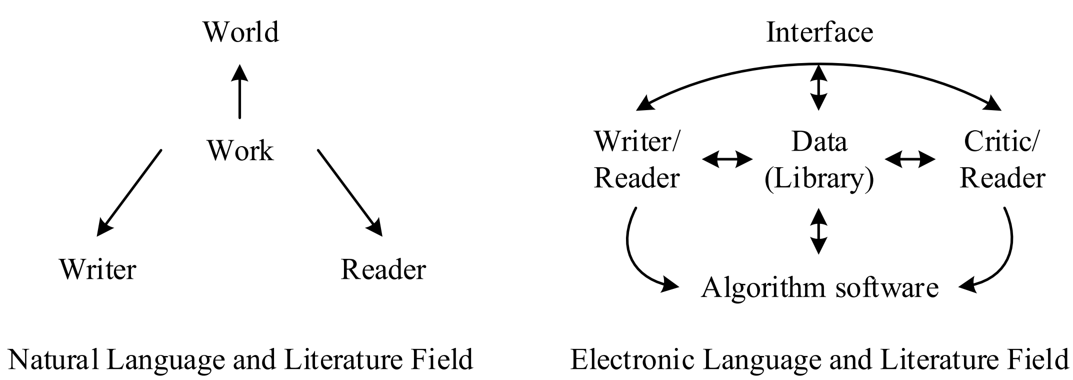


Figure 1: The five-element theory of digital humanistic criticism in literature

Currently in the new stage of the establishment of the Large Language Model (LLM), the narrative rules, patterns and their methods of open source literary knowledge production have become the focus of research. Utilizing the generative AI algorithm software based on the Large Language Model and other digital technologies, methods, and tools, the exploration and practice paths of AI-generative literature in future literary research are as follows:

(1) The digital humanities critical discourse of literature is being formed. Since the digital humanistic criticism of literature is based on literary and digital symbols, emphasizing the re-expression of literary critical discourse on the basis of digital narrative, digital tool discourse and technological methodological terminology, the discourse mechanism of “Intertextual Expression” has been formed, which provides a basis for the formation of the discourse of A1-generative literary narrative and ethical criticism.

(2) To create a specialized generative artificial intelligence big language model of “literary criticism” through literary modeling, firstly, digitize the collected paper texts and store them in an online dynamically updated corpus. Secondly, the data collection technology is used to aggregate the literary research literature that already exists in the virtual world according to disciplines, and then continue to store it in the same online dynamically updated corpus. Finally, these digitized and organized texts are input into the AI digital human literary critic model, trained based on the deep learning technology of interpretable artificial intelligence, to create a generative AI digital human literary critic with algorithmic logic, which interacts and dialogues with human critics, and realizes a complete interpretation of the literary research object.

(3) Explore the establishment of a literary evaluation system adapted to generative AI literary creation, criticism, and AI generative literary research. For example, to establish AI narratology, to think about the differences between AI narrative perspectives, identities, and values and human narrative perspectives, identities, and values, how to avoid problems such as prejudice and discrimination that may be brought about by digital technology, and how to utilize digital technology to balance the aesthetic value and social responsibility of literary works. Another example is the difference between literary criticism on digital platforms and the criticism of critics in the past, which involves the digitization and multimedia of literary research and dissemination of results, with a large number of interactive reviews flooding various video platforms, and such results integrating presentations, reviews, and audience interactions, which can be open-source shared resources or treatises. Such results retain the live sense of critical dialog and support subsequent researchers to participate in the discussion through pop-ups, forming a continuously updated academic dialogue. However, it also poses challenges such as quality control, copyright and academic authentication.

In conclusion, in the context of digital intelligence, the way of textual existence, writer's identity, reader/critic's identity, and their interactions in traditional literature have undergone qualitative changes, producing digital texts, generative AI texts, and pluralistic differences between AI identities and creative roles. These changes call for a new theoretical interpretation of digital humanities criticism of literature.

III. Sentiment Analysis Based on RoBERTa-BiLSTM-Attention

In order to realize AI-based generation of ethical criticism of digital literary narratives, it is necessary to calculate the emotion of the text of literary works, for this reason, this paper proposes a method of sentiment analysis based on the RoBERTa-BiLSTM-Attention model. The goal of sentiment analysis is to classify the text based on its emotional polarity, which is based on the emotions expressed in the text, which can be mainly categorized into different types of emotions such as positive, negative and neutral.

III. A. Relevant Technology and Theoretical Foundations

III. A. 1) Text pre-processing

Text preprocessing in natural language processing is a crucial step, which includes disambiguation processing, deletion of deactivated words, text serialization and padding processing, etc. The following will elaborate on disambiguation and disambiguation processing.

(1) Deactivation

Pre-processing text, often remove the disabling words. Disposable words, that is, words that do not contain specific meanings, usually play a role in connecting sentence components or complementary role, and removing them will not affect the expression of sentiment in the sentence. Therefore, removing these stop words is particularly important when conducting sentiment analysis. In addition, removing stop words can also reduce the spatial dimension of data features and improve the efficiency of model training.

(2) Segmentation

In Chinese text, Chinese characters are used as the basic unit, and a series of Chinese characters are used to represent a series of words. However, there is no clear distinction between these words, so specialized algorithms must be used to differentiate them. Currently, common Chinese word separation algorithms include: string matching method, probability statistics method, and semantic understanding method.

III. A. 2) Text word vectorization

(1) One-hot Model

One-hot coding uses N-bit state registers to describe multiple states, and it converts categorical variables into binary form to better understand and describe complex data. Firstly, a dictionary is constructed based on the text, in which the numbers can be regarded as the labeling information of the corresponding words or the categorization information of the transactions. Then an N-dimensional vector is constructed based on the uniquely hot coded expression, the dimension of the vector is always the same as the length of the dictionary, and for the vector expression of a given word, the register of the response position of its occurrence in the dictionary is assigned as 1, and the rest is 0.

(2) TF-IDF algorithm

TF-IDF is an effective technique for word frequency analysis, which can help us to identify the keywords in the article more precisely and reflect the importance of certain words in the article by calculating the inverse text frequency index IDF and word frequency TF [16].

The main implementation steps of TF-IDF algorithm are as follows:

Step1: Calculate the word frequency TF:

$$\text{Word frequency (TF)= the number of times a certain word appears in the text} \tag{1}$$

Due to the differences between the texts, word frequency harmonization needs to be carried out first in order to facilitate the comparison of different texts. In the process of word frequency harmonization, the vocabulary in each text needs to be counted and calculated, and then transformed into a standard format, and the alternative formula for calculating word frequency is:

$$\text{TF} = \frac{\text{The number of times a certain word appears}}{\text{The occurrence frequency of the word that appears most frequently}} \tag{2}$$

Step2: Firstly, inverse text frequency (IDF) is calculated. By utilizing the corpus, it is possible to capture and describe the performance of the language in different scenarios in a finer way, and the calculation formula is as follows:

$$\text{IDF}=\log\left(\frac{\text{The total number of texts in the corpus}}{\text{The number of texts containing this word} + 1}\right) \tag{3}$$

As the frequency of a word increases, its inverse document frequency decreases and eventually approaches 0. In order to prevent all texts from not containing the word, it is inherent to add 1 to its denominator. In addition, by using a logarithmic function, the denominator can be made larger to more accurately reflect the word's importance in the text.

Step3: Calculate the TF-IDF:

$$\text{TF-IDF}= \text{TF} \times \text{IDF} \tag{4}$$

When the TF-IDF value of a word is close to 1, it indicates that the word has a more important position in the text.

(3) Doc2vec model

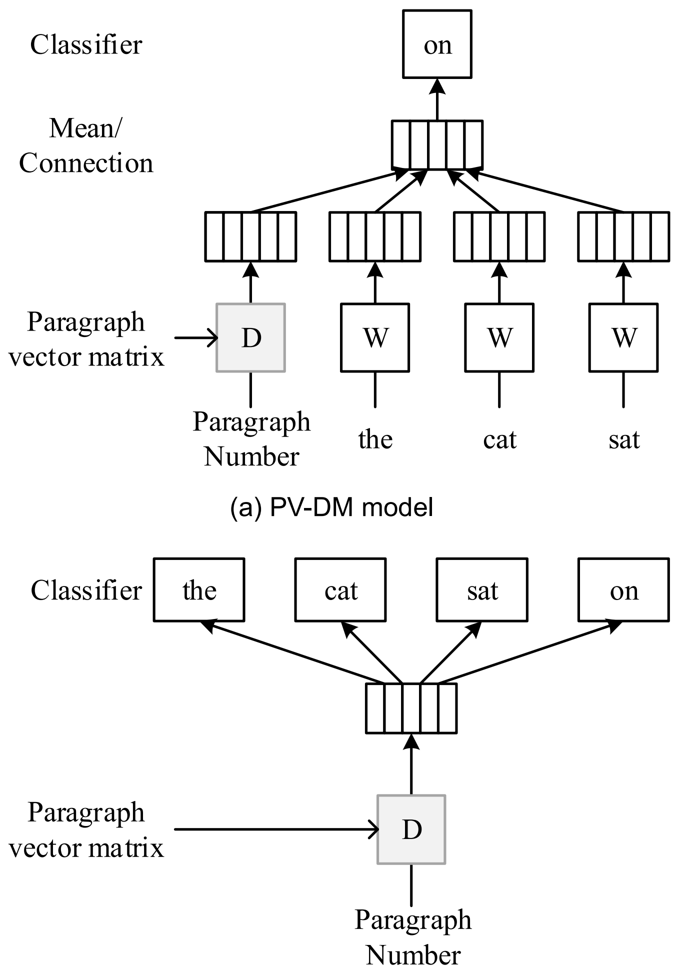
Doc2vec model [17], also known as Paragraph Vector, which is designed to create a vectorized representation of a document without the limitation of its length.Doc2vec model is an extension of Word2vec.Doc2vec embeds word vectors while generating an additional vector for the entire document to capture the semantic information of the entire text file. The model can be categorized into two algorithms: paragraph vectors with distributed memory (PV-DM) and paragraph vectors with distributed bag-of-words version (PV-DBOW).

The structure of the two Doc2vec models is shown in Fig. 2, where (a) denotes the PV-DM model and (b) denotes the PV-DBOW model. The PV-DM model achieves effective identification of the features of each paragraph by mapping it to a specific Paragraph id, which improves the accuracy of the model. Each sentence is converted into a unique vector as a column of a matrix. Each word is converted into a unique vector, also as a column in a matrix. By combining the sentence vectors and all the word vectors into one matrix, the next possible occurrence of a word can be calculated. For this purpose, sentence vectors and word vectors are averaged or concatenated to predict the next word in the text. In contrast, the PV-DBOW model is characterized by the fact that it does not depend on the context, but takes a unique approach where key words are randomly extracted from the whole sentence for prediction.

(4) BERT model

BERT is a variant of Transformer, a self-encoder language model. The process of BERT pre-training is to gradually adjust the model parameters so that the semantic representation of the text output by the model can portray the nature of the language, which facilitates the subsequent fine-tuning for the specific Next Sentence Prediction (NSP) task. In order to achieve this, the BERT model can be greatly improved by performing two predefined tasks, Masked Language Modeling (MLM) as well as NSP, which can better fulfill diverse natural language processing tasks.

1) The first task aims to train the language model using the MLM approach. Specifically, when a sentence is input, some words to be predicted are randomly selected and special symbols [MASK] are used to replace them. Then, the model is allowed to learn which words should be filled in at these positions based on the labels, thus improving its language comprehension.



(a) PV-DM model

(b) PV-DBOW model

Figure 2: Structure of the Doc2vec models

2) The second task aims to extend the bidirectional language model by introducing a new technique for examining whether there is an obvious connection between two pieces of text in an article, and whether there is some degree of interdependence between them.

The BERT model is trained jointly on the MLM task and the NSP task, so that the vector representation of each word and phrase output from the model can portray the overall information of the input text as comprehensively and accurately as possible. The input structure of BERT is shown in Fig. 3, which consists of the positional encoding of the current word, the sentence encoding of the sentence in which the current word is located, the current word encoding, and the sentence beginning of the first sentence is encoded with [CLS] to identify it, and different sentences are identified by [SEP].

III. A. 3) Attention mechanisms

Attention mechanism is a way of information filtering, which can effectively alleviate the long-time dependency problem and is widely used in LSTM and GRU. Attention mechanism can be used in two ways: independently or as a layer of the hybrid model, which can be placed behind the text vector input layer or other network models. This mechanism links different parts of the model through automatic weighted transformations to highlight key words and make the whole model perform better. The structure of the attention mechanism is shown in Figure 4.

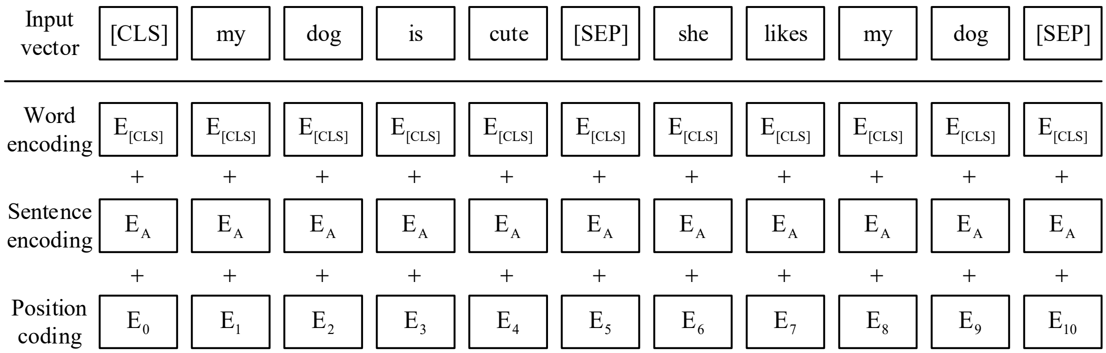


Figure 3: The input of BERT

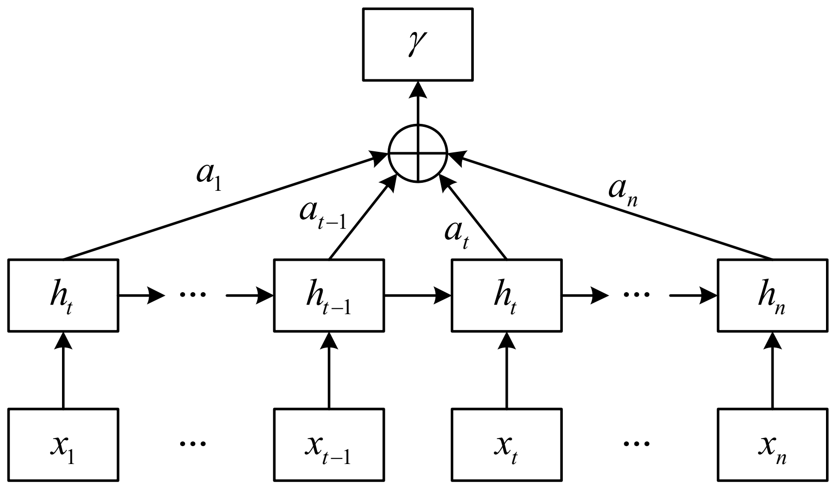


Figure 4: Basic Structure of the Attention mechanism

III. B. RoBERTa pre-training model

RoBERTa [18] is optimized on the basis of BERT model [19], which is a robust and optimized pre-training method for BERT. The overall structure of the model is basically the same as that of BERT, but there are large differences in training methods, hyperparameter settings, etc. It can be applied to more difficult tasks with higher generalization ability, and achieves better results than the BERT model on many tasks.

The RoBERTa model has the following main improvements:

(1) Dynamic masking mechanism is used. The BERT model performs random MASK just once for each sample in the data preprocessing stage, and the order of word MASK is the same for each round of training after that, which is a static MASK operation. RoBERTa, on the other hand, uses a dynamic masking mechanism, where a new masking pattern is generated each time a sequence is input to the model. In this way, the model gradually adapts to different masking strategies and learns different linguistic representations as a large amount of data is continuously input.

(2) The NSP task in BERT is removed and a larger byte-level BPE vocabulary is used to train the model.

(3) Increased Batch size. RoBERTa uses a larger bacth size during training. Tried bacth sizes ranging from 256 to 8000.

(4) Increased training data. In BERT, Book Corpus and Wikipedia datasets were used for training. RoBERTa, on the other hand, uses more textual data, including CC-NEWS, Stories, and other datasets, totaling over 160 GB of textual data. This enables RoBERTa to learn a wider range of language knowledge.

III. C. RoBERTa-BiLSTM-Attention modeling

Based on the RoBERTa-BiLSTM-Attention sentiment analysis method, firstly, the text is inputted into the RoBERTa model for word vectorization to get the sentence embedding and word embedding of the text, then the word embedding is used to learn the contextual features through the bi-directional long- and short-term memory network, and the attention is focused on the more important information in the text through the attention mechanism, and the output obtained is spliced with the sentence embeddings are spliced, the input to the fully connected layer maps the feature representation to the sentiment label space, and finally the output of the fully connected layer is normalized using the Softmax function to obtain the classification results. The overall structure of the sentiment analysis model based on RoBERTa-BiLSTM-Attention is shown in Fig. 5, and it can be seen that the model consists of a text vector representation layer, a feature extraction layer and an output layer.

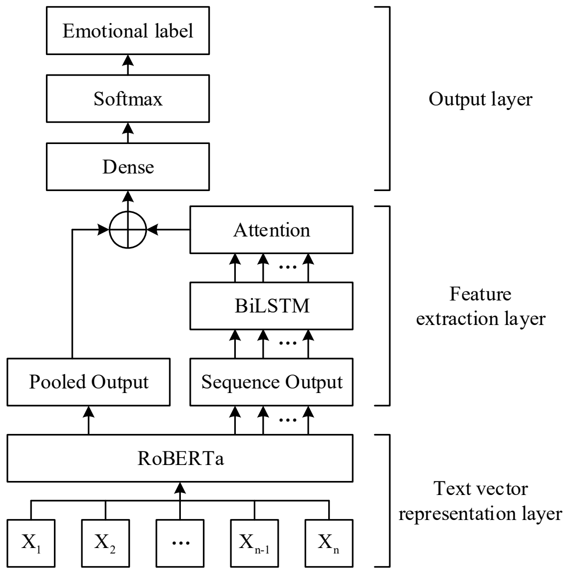


Figure 5: The structure of RoBERTa-BiLSTM-Attention model

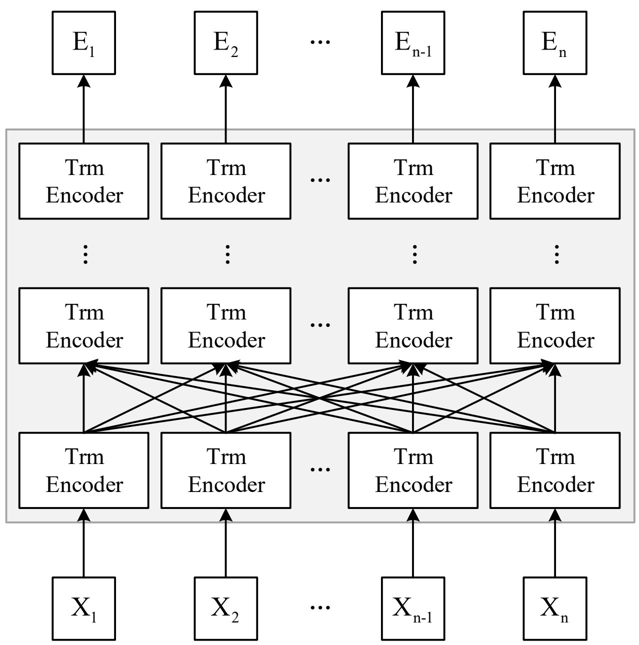


Figure 6: RoBERTa layer structure

III. C. 1) Text Vector Representation Layer

The text vector representation layer mainly converts the input text into the form of word vectors. In this paper, RoBERTa pre-trained language model is used to generate the word vector representation of the input text. It is assumed that the input $X=[X_I,X_2,\ldots,X_N]$, N refers to the total length of the input, and the vector representation of each word is a superposition of word embeddings, sentence embeddings, and positional embeddings. Input X into the RoBERTa model for word vectorization, the vectorized text data is $E=[E_1,E_2,\ldots,E_N]$, and the structure of the RoBERTa layer is shown in Figure 6.

After the word vectorization of RoBERTa model, two kinds of outputs can be obtained, one is the output of the identifier $[\mathrm{CLS}]$ in the last layer, because this symbol with no obvious semantic information will be more “fair” to integrate the semantic information of the words in the text compared with other words in the text, and so it can better express the semantic information of the whole sentence, which contains the overall characteristics of the text, and can be used as the sentence embedding of the input text as a whole, i.e. the Pooled Output part of the figure. The semantic information of the whole sentence, including the overall characteristics of the text, can be used as the overall sentence embedding of the input text, i.e., the Pooled Output part of the figure, which is denoted by C, and there is another kind of output of all the words of the input text at the last level, which can be used as the word embedding of the text, i.e., the Sequence Output part of the figure, which is denoted by S.

III. C. 2) Feature extraction layer

The feature extraction layer consists of the Bidirectional Long Short-Term Memory Network (BiLSTM) [20] layer and the Attention Mechanism layer together, firstly, the word embedding S of each word output from the Text Vector Representation layer is inputted into the BiLSTM layer, because BiLSTM combines forward LSTM and reverse LSTM, it can better obtain the bidirectional semantic features of the input information, so as to effectively extract the input text the emotional information contained in the input text. At any moment, both forward and reverse LSTMs get the input information at the same time. When extracting semantic features, the forward LSTM computes the input S_t at moment t and the output O_{t-1} at moment $t-1$ to get the forward output O_t at moment t, which is computed as follows:

$$O_t = f(W_1 S_t + W_2 O_{t-1}) \tag{5}$$

where W_1, W_2 are the weights in the computation process.

The reverse LSTM computes the input S_t at moment t with the output P_{t+1} at moment $t+1$ to obtain the reverse output P_t at moment t, which is computed as follows:

$$P_t = f(W_3 S_t + W_4 P_{t+1}) \tag{6}$$

where W_3, W_4 are the weights in the computation process.

The hidden layer then combines the outputs of both directions to finally compute the output H_t of the BiLSTM at moment t, which is computed as follows:

$$H_t = g(W_s O_t + W_6 P_t) \tag{7}$$

where W_5, W_6 are the weights in the calculation process.

Next, the results of BiLSTM are inputted into the attention mechanism layer, which assigns different feature weights according to the different importance of the words in the text, better learns the semantic relationship between words, improves the attention to the important words, and reduces the noise in the text, so as to further improve the accuracy of sentiment classification.

The attention mechanism layer first multiplies the output H of BiLSTM by the weight parameter W_c, plus the bias b_c of the attention mechanism layer, and then nonlinearizes it with the tanh activation function, which is calculated as follows:

$$F = \tanh(W_c \cdot H + b_c) \tag{8}$$

The obtained F is normalized with Softmax function to obtain the weight value α, which is calculated as follows:

$$\alpha = Softmax(F) \tag{9}$$

The weighted sum is obtained by multiplying the weight value α with the output H of BiLSTM to get the final output R, which is calculated as follows:

$$R = \sum \alpha \cdot H \tag{10}$$

Finally, T is obtained by splicing the output R of the Attention Mechanism layer with the sentence embedding C of the text output from the Text Vector Representation layer, which is formulated as follows:

$$T = [R, C] \tag{11}$$

III. C. 3) Output layer

The output layer consists of a fully connected layer and a Softmax function, while in order to prevent overfitting problems in the fully connected layer to join the Dropout mechanism, through the use of the Dropout mechanism can be randomly discarded neurons in the neural network, thus ensuring that the training of the neural network to become more in-depth, so that the trained model has a stronger ability to generalize, and does not have an over-reliance on some local features. The working schematic of Dropout is shown in Fig. 7:

Where, the left figure shows the neural network training process without using Dropout, while the right figure shows the case where Dropout is used. After the Dropout mechanism is added to the fully connected layer, the formula is as follows:

$$y = Dropout(W_d T + b_d) \tag{12}$$

where W_d is the weight value of the fully connected layer and b_d is the bias.

Finally, the Softmax function is used for normalization, and each value in y is mapped to between 0 and 1 to get the probability value Y of different emotional tendencies, where the maximum value corresponds to the category that is the emotional tendency of the input text. The formula is as follows:

$$Y = Softmax(y_i) = \frac{\exp(y_i)}{\sum_j \exp(y_j)} \tag{13}$$

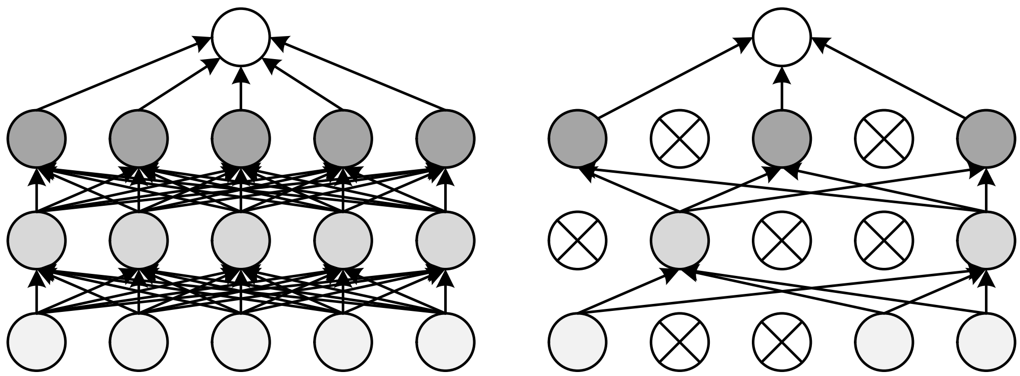


Figure 7: Illustration of Dropout

III. D. Model experiment and result analysis

In order to verify the specific effect of the proposed RoBERTa-BiLSTM-Attention model in the sentiment analysis of literary texts, this section conducts application experiments on the model.

III. D. 1) Experimental dataset construction

The domain selected for this experiment is mainly the domain of literary reviews. The experimental data of this experiment is divided into two parts, the first part is a sentiment polarity dataset A about literary reviews, which contains a total of 8,045 literary reviews that are labeled as positive or negative sentiment polarity. Among them, there are 5463 positive reviews and 2582 negative reviews. The second part is the dataset B of reviews of digital literary works crawled using crawling techniques, which collects a total of 72,685 literary review data, including 62,924 positive reviews and 9,761 negative reviews. All data samples are labeled by manual labeling, where positive samples are labeled as 1 and negative samples are labeled as 0.

Finally, the two parts of the dataset are combined together and used as the training set, validation set and test set in the ratio of 8:1:1, respectively, and the specific data volume distribution is shown in Table 1.

Table 1: Dataset data

Name	Positive sample size	Negative sample size	Total
Training set	54709	9875	64584
Verification set	6839	1234	8073
Test set	6839	1234	8073

III. D. 2) Experimental evaluation indicators

For the experimental evaluation of the validation set, Accuracy (Acc), Precision (P), Recall (R) and the combined evaluation metric F1 value were used to measure the performance of the model. The formula is calculated as follows:

$$Acc = \frac{TP + TN}{n} \tag{14}$$

$$P = \frac{TP}{TP + FP} \tag{15}$$

$$R = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{17}$$

where the true example TP, false positive example FP and false negative example FN denote the classification results of the model, while the total number of samples n denotes the total number of samples in the validation set.

III. D. 3) Experimental Environment and Parameter Configuration

(1) Experimental environment

This experiment is based on the PyTorch framework, which involves the configuration of the parameters as shown in Table 2.

Table 2: Experimental environment parameters

Category	Parameter setting
Operating system	Ubuntu 18.04.2LTS
CPU	Inter(R)Xeon(R)Gold 5218R
GPU	RTX 3060
Memory	64G
Python version	Python3.13
PyTorch version	Py Torch2.7

(2) Parameter settings

The experimental hyper-parameter configuration is shown in Table 3.

Table 3: Experimental hyperparameter configuration

Parameters	Value
Pre-trained model	RoBERTa-wwm-ext
Maximum text length	128
Learning rate	1e-5
Discard rate	0.1
Batch processing size	64
Optimizer	Adam

III. D. 4) Experimental results and analysis

(1) Ablation experiment and analysis

In order to verify the effect of the multi-head attention mechanism layer in the model proposed in this paper on the effect of analyzing the textual sentiment of literary works, ablation experiments were designed on the dataset for comparison. These comparison models include the model of this paper with the multi-head attention mechanism layer and the RoBERTa-BiLSTM model without the multi-head attention mechanism layer. The results of the ablation experiments are shown in Figure 8.

The data show that this paper's model with a multi-head attention mechanism layer improves 3.22% and 1.65% in the two evaluation metrics, accuracy and F1, respectively, over the RoBERTa-BiLSTM model without a multi-head attention mechanism layer. The difference between the two models in the evaluation metrics is mainly due to the fact that although the BiLSTM model can capture the contextual information and global features of words, it cannot effectively capture the semantic relationships between words. In contrast, the multi-attention mechanism is able to pay different degrees of attention to different parts of the input, including different positions and semantic perspectives, thus better capturing the semantic relationships between words and helping to improve the effectiveness of text sentiment analysis.

(2) Comparative Experiments and Analysis

In order to further prove the effectiveness and accuracy of the sentiment analysis model based on RoBERTa-BiLSTM-Attention, this paper, in the same experimental environment, compares it with five other sentiment analysis models such as Word2vec-BiLSTM, BERT-BiLSTM, RoBERTa-BiLSTM, Word2vec-BiLSTM-Attention and BERT-BiLSTM-Attention models. Attention and BERT-BiLSTM-Attention models, and five sentiment analysis models were studied against each other, considering accuracy and F1 value as the main evaluation metrics. All these models were trained on the same set of training data and then underwent testing on a unified set of test data, and the results of the model comparison experiment are shown in Table 4.

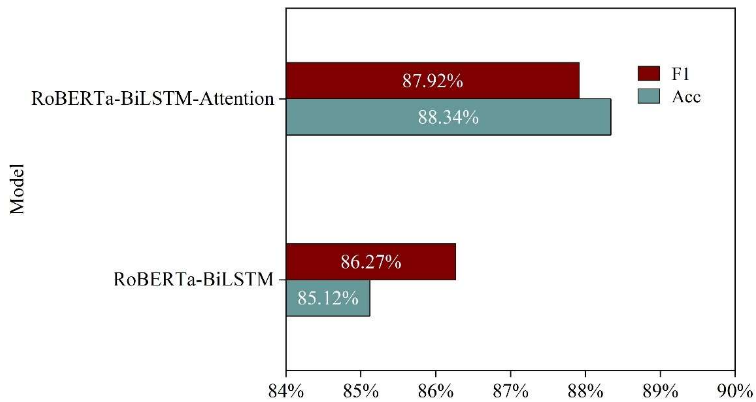


Figure 8: Comparison of models with and without multi-head attention mechanism layers

Within the framework of this comparison experiment, because the conventional Word2vec method used by the Word2vec-BiLSTM model cannot deal with lexical polysemy, it slightly underperforms among all the tested models. However, this model still serves as an important reference point for emotion recognition model improvement and helps to provide insight into how various factors affect the performance of emotion recognition models.

The BERT-BiLSTM model combines the powerful feature extraction capability of the BERT model with the advantages of a bidirectional LSTM network, resulting in a better ability to capture deep linguistic traits. Compared to the ordinary Word2vec-BiLSTM model, the BERT-BiLSTM model performs better in sentiment classification, with significantly higher accuracy and F1 scores than the former. And the BERT-BiLSTM-Attention model is based on BERT-BiLSTM with the addition of the attention mechanism, which is able to adjust the importance of the feature information according to their weights, which in turn improves the effect of text emotion recognition.

The model using RoBERTa-BiLSTM-Attention successfully replaces BERT with RoBERTa and effectively combines the word embeddings and sentence embeddings produced by the pre-trained model, which enables the model to acquire richer semantic information. It can be seen that the model of RoBERTa-BiLSTM-Attention performs the best in comparing the various models in terms of both accuracy and F1 value, reaching 88.35% and 86.52% respectively, which is superior to the performance of all other models. This indicates that the model in this paper is suitable for the task of recognizing the emotions of literary texts and can be used for AI to generate ethical criticism of literary narratives.

Table 4: Comparison of experimental results

Models	Acc/%	P/%	R/%	F1/%
Word2vec-BiLSTM	82.19	83.25	78.64	80.85
BERT-BiLSTM	85.74	87.51	82.38	84.87
RoBERTa-BiLSTM	87.26	87.84	82.42	85.04
Word2vec-BiLSTM-Attention	83.53	84.36	79.57	81.90
BERT-BiLSTM-Attention	86.12	88.45	83.25	85.77
RoBERTa-BiLSTM-Attention	88.35	89.12	84.06	86.52

IV. Criticism of Narrative Ethics in Chinese New Century Literature Based on Emotional Analysis

In this chapter, author Su Tong's long story “Riverbank”, a Chinese new century literary work, is selected as a specific research object, and the constructed RoBERTa-BiLSTM-Attention model is used to conduct a textual sentiment analysis, which provides a basis for generating ethical criticism of literary narratives using AI.

IV. A. Sentiment analysis of text based on trend curves

The novel text of The Bank of the River has a strong and complex lyrical color, and through the perspective of Ku Dongliang, it deeply digs into the past of his father, Ku Wenxuan. The text incorporates a rich combination of emotions, and the semantic expression is compact and clear, so special attention needs to be paid to the relative relationship between emotional features when analyzing it. Text sentiment analysis refers to the process of analyzing, processing, summarizing and extracting the subjective sentiment of a text using techniques such as natural language processing and text mining. Using ROST EA software, the text of the novel was emotionally subcalculated and coded to form a text-emotion time series, followed by the use of nonlinear adaptive filtering (NAF) method to obtain the emotional trend line of the script, and the development of the emotional trend of the novel “Riverbank” was portrayed using quantitative methods as shown in Figure 9.

The emotional trend line of the text of The Riverbank basically fluctuates in the negative interval, which is consistent with the negative emotional tone of the text of anger and fear. The direction of fluctuation is largely positive, consistent with the emotional state of the text, which is intertwined with hope and despair. The emotional trend line drops sharply in the negative range in the early part of the text, as the father flees to the Yangtze River with his son, the mother attempts suicide, and the family ethics are completely torn apart from the political oppression. The emotional trendline then fluctuates upward, reflecting the brief freedom of the father and son during their wanderings on the river, with positive segments such as the old boatman's help alternating with negative existential oppression, and a mixture of novelty, loneliness, hope and despair. Subsequently, the father is arrested and paraded through the streets, and the emotional trend line falls off a cliff to its lowest point and reaches a negative peak, corresponding to the emotional keywords of humiliation, anger, and powerlessness to resist. Finally, the father is escorted back to the town of Oil Mill, and the people around him are implicitly sympathetic in their silence, and the affective trend line slowly rises back up, reflecting the individual awakening and goodwill of the people. Positive emotions are implied in negative emotions, and the ending text tends to be emotionally neutral, with complexity implied in the white space, suggesting that the cycle of history has not ended.

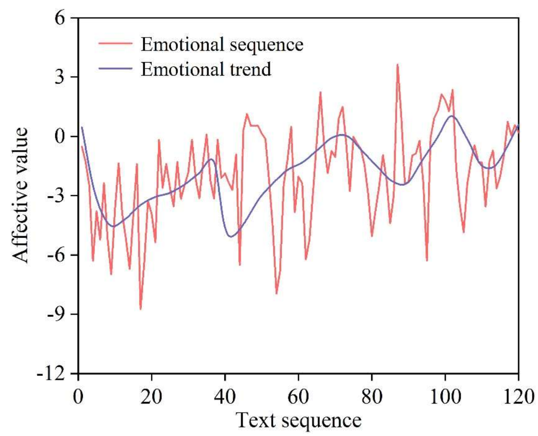
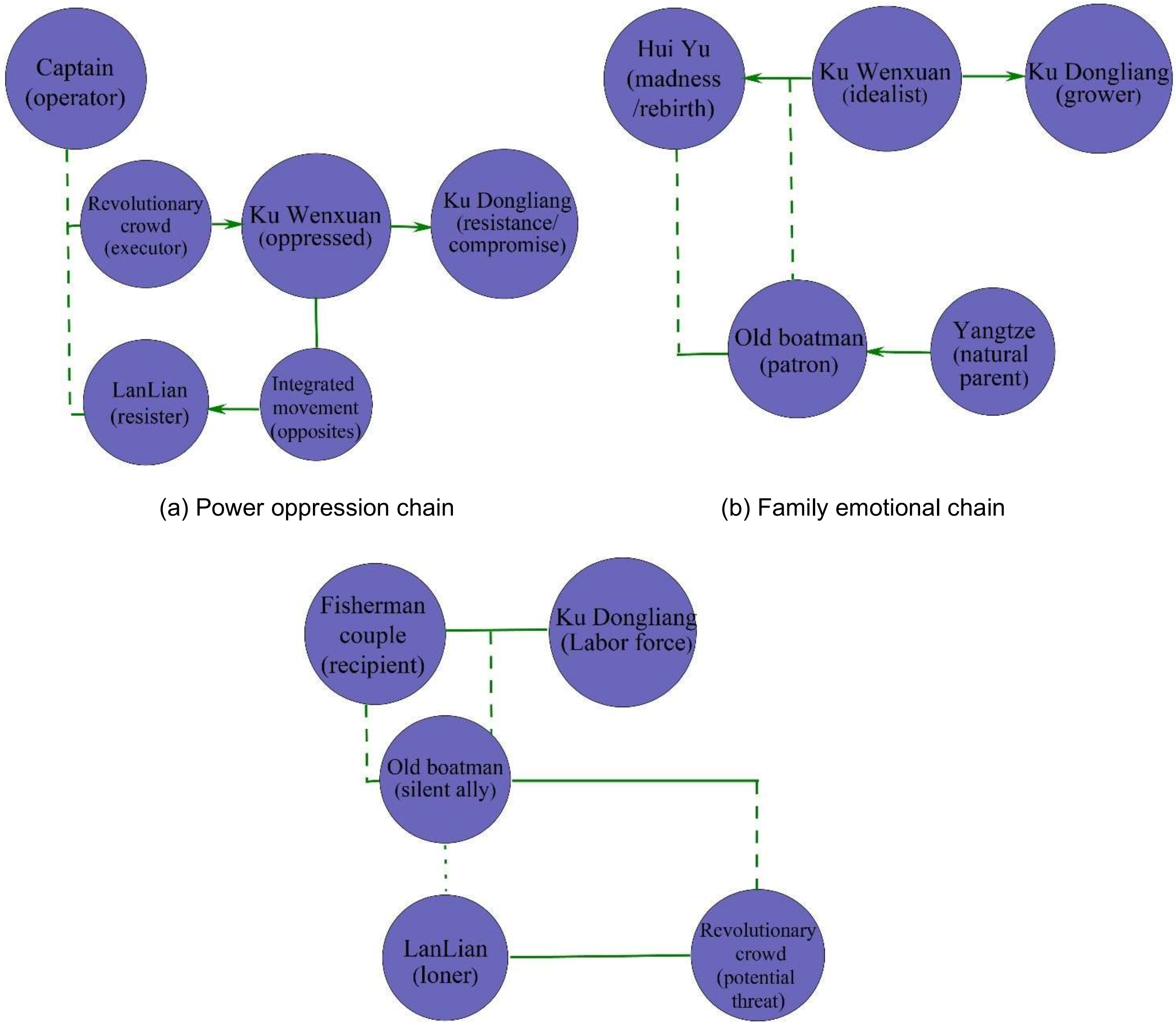


Figure 9: Text Sentiment Trend

IV. B. Analysis of Character Role Relationships Based on Social Networks

Characters are the central elements of the novel, and the key to understanding the plot and setting of The Riverbank lies in recognizing the importance of characters and sorting out their relationship lines. Differentiation. Social network analysis is an analytical method used to view the interrelationships between nodes and connecting edges, which can be used to measure the importance and connections of each participant in the text. The relationships of the characters in the text of The Riverbank are extracted from the text and the visualization of the characters' social network relationships is shown in Figure 10, where (a) ~ (c) denote the chain of power oppression, the chain of family emotions, and the chain of underlying dependence, respectively. Taking the character roles in the text as nodes, connecting nodes through edges and labeling the types of relationships between different roles in parentheses. The larger a node is, the more important the character is in the play, and the more relationship edges a character has means that its association with other characters in the novel is more complex. In addition, arbitrarily clicking on a character's avatar provides further insight into their character lineage and relationships in the text.

The character network of The Riverbank is dispersed with Ku Wenxuan as the core character, and all the characters except Blueface are directly connected to Ku Wenxuan and narrate the character relationships through Ku Dongliang's point of view, thus forming a more centralized character network diagram.



(a) Power oppression chain

(b) Family emotional chain

(c) Underlying dependency chain

Figure 10: Network relationship of character roles

IV. C. Word cloud-based high-frequency word analysis

The word cloud is a visual presentation of the high frequency “keywords” in the text, which conveys the valuable information behind a large amount of text data through the combination of text, color and graphics. The word cloud of high-frequency words in the text of Riverbank is shown in Figure 11.

From the word cloud, it can be clearly seen that the high-frequency words in the text of "Riverbank" are related to the names of the characters such as Ku Wenxuan, Ku Dongliang, Lao Gonggong and other core characters, which shows that the protagonist of the novel occupies a core position in the text as the main body of plot promotion and emotional expression. In terms of natural imagery, the author uses "rivers", "river banks", and "fish" to reflect the dissolution and contrast of nature on political violence. In terms of political violence, it shows how the system alienates people into symbols through "rightists", "criticism", and "labor reform". In terms of character state, psychological words such as "silence", "madness" and "fear" are used to reflect the individual's spiritual collapse and survival strategy under oppression. In terms of survival actions, the essence and struggle of marginalized survival are reflected through "escape", "labor", and "hiding".

Figure 11: High word cloud

V. Conclusion

This paper proposes a sentiment analysis method based on the RoBERTa-BiLSTM-Attention model, and uses AI to generate ethical criticism of literary narratives on this basis.

Compared to the RoBERTa-BiLSTM model without adding the multi-head attention mechanism layer, the model in this paper improves 3.22% and 1.65% in accuracy and F1 metrics, respectively, reflecting the utility of the attention mechanism to improve the performance of the model. Meanwhile, comparing the five models, including Word2vec-BiLSTM, BERT-BiLSTM, RoBERTa-BiLSTM, Word2vec-BiLSTM-Attention, and BERT-BiLSTM-Attention, this paper's model performs the best in terms of both accuracy and F1 value, which are as high as 88.35% and 86.52% respectively. This indicates that the organic combination of the three parts of the model can realize the efficient analysis of the emotions of the literary text, and can provide the basis for the AI to generate the ethical criticism of the literary narrative.

Taking the text of the novel Riverbank as an example, we analyze the change of emotional trend of Chinese literature in the new century, and the results show that the emotional trend curve basically fluctuates in the negative range, which is consistent with the negative emotional tone of anger and fear in the text. The direction of fluctuation is largely positive, which is consistent with the emotional state of hope and despair intertwined in the text. The character network of The Riverbank is dispersed with Ku Wenxuan as the core character, and the character relationships are narrated through Ku Dongliang's perspective, thus forming a more centralized character network diagram. In addition, the results of high-frequency word crawling show that Riverbank embodies the process and emotional changes of the characters' struggle against political oppression in terms of natural imagery, political violence, character states, and in survival actions.

Funding

Zhuhai philosophy and social science research ' Zhuhai leisure culture research base ' project results.

References

[1] Jiang, Z. (2015). Identification of Issues Concerning Contemporary Western Literary Criticism: With Concurrent Reflections on the Reconstruction of Chinese Literary Criticism. Social Sciences in China, 36(1), 5-29.
[2] Shunqing, C. (2008). The discourse of Chinese literary theory and the dialogue between Western and Chinese literary theories. Journal of Multicultural Discourses, 3(1), 1-15.
[3] Zeng, J. (2020). Rethinking Chinese Questions in Western Literary Theories. Critical Arts, 34(2), 1-4.
[4] Liu, X. (2023, February). Jia Pingwa's Prose View. In 2022 4th International Conference on Literature, Art and Human Development (ICLAHD 2022) (pp. 963-971). Atlantis Press.
[5] Wang, B., & Barr, A. H. (2021). Yu Hua's works in English translation: An interview with Allan H. Barr. Asia Pacific Translation and Intercultural Studies, 8(1), 84-97.
[6] Qinghua, Z., & Lingenfelter, A. (2013). The Nobel Prize, Mo Yan, and Contemporary Literature in China. Chinese Literature Today, 3(1-2), 17-20.
[7] Gaffric, G., & Peyton, W. (2019). Liu Cixin's Three-Body Trilogy and the Status of Science Fiction in Contemporary China. Science Fiction Studies, 46(1), 21-38.
[8] Youquan, O. (2011). Chinese literature's transformation and digital existence in the new century. Social Sciences in China, 32(1), 146-165.
[9] Chen, X., & Sheng, A. (2013). A Survey of Twentieth-century Literary Theory and Criticism in Chinese. CLCWeb: Comparative Literature and Culture, 15(6), 14.
[10] Wang, C. M. (2012). Geopolitics of literature: Foreign literature studies in early twentieth-century China. Cultural Studies, 26(5), 740-764.

[11] Zheng-shuan, L. I., & Hui-min, Z. H. U. (2022). The Latest Development of Ethical Literary Criticism in the World. Journal of Literature and Art Studies, 12(1), 1-11.
[12] Ning, W. (2013). Globalizing Chinese Literature: toward a rewriting of contemporary Chinese literary culture. In China's literary and cultural scenes at the turn of the 21st century (pp. 103-118). Routledge.
[13] Yang, Y. (2022). Contemporary full-length novels: A literary symbol. In Chinese social sciences and humanities studies: Collection of important topics (pp. 17-29).
[14] Hu, Y. (2023). Nation: A New Dimension of Literary Criticism in the Context of Globalization. In The Contemporary Construction of the Chinese Form of Marxist Literary Criticism (pp. 51-87). Singapore: Springer Nature Singapore.
[15] Longxi, Z. (2018). World literature, canon, and literary criticism. Tensions in world literature: Between the local and the universal, 171-190.
[16] Seonghyun Park, Seungmin Oh & Woncheol Park. (2025). Automated Classification Model for Elementary Mathematics Diagnostic Assessment Data Based on TF-IDF and XGBoost. Applied Sciences,15(7),3764-3764.
[17] Qiang Gao & Man Jiang. (2024). Exploring technology fusion by combining latent Dirichlet allocation with Doc2vec: a case of digital medicine and machine learning. Scientometrics,129(7),4043-4070.
[18] Biplov Paneru, Bipul Thapa & Bishwash Paneru. (2025). Sentiment analysis of movie reviews: A flask application using CNN with RoBERTa embeddings. Systems and Soft Computing,7,200192-200192.
[19] Gaoqing Xu, Qun Chen, Shuhang Jiang, Xiaohang Fu, Yiwei Wang & Qingchun Jiao. (2025). Analyzing the capability description of testing institution in Chinese phrase using a joint approach of semi-supervised K-Means clustering and BERT. Scientific Reports,15(1),11331-11331.
[20] Jianwei Liang, Liying Yin, Yanli Xin, Sichao Li, Yuqian Zhao & Tian Song. (2025). Short-term photovoltaic power prediction based on CEEMDAN-PE and BiLSTM neural network. Electric Power Systems Research,246,111706-111706.