Depression Detection Based on Multi-source Information Fusion

Weifeng Yuan *

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, China

* Corresponding Author Email: seanight_1@outlook.com

Abstract. In recent years, depression has severely impacted the physical and mental health and safety of adolescents, making both treatment and prevention of depression in adolescents crucial. To enhance depression detection, this paper proposes a depression detection model based on the RBF kernel and the BERT Transformer. This model extracts features from text data and structured feature data using a text feature extraction module and a structured data extraction module, respectively. After feature fusion, it uses support vector machines for classification and finally calculates the probability of depression and the depression index of the student. Experimental validation demonstrates excellent results on a student depression dataset containing 27,902 records, achieving an AUC of 1.00. The weighted depression dictionary fusion method increases the sensitivity of high-risk semantic features, facilitating depression detection. Compared to traditional models, the RBF kernel and the BERT Transformer provide a superior solution for building depression detection systems, contributing to enhanced mental health protection and self-defense awareness.

Keywords: BERT Transformer, TF-IDF, Depression Dictionary, SVM, Depression test.

1. Introduction

Adolescent depression is a major social and public health problem. The prevalence of depression is relatively high, usually accounting for more than 10% of patients [1]. Adolescents with depression will experience symptoms of drowsiness, depression, and psychosis [2]. Therefore, it is particularly important to detect adolescent depression and provide early intervention and treatment. Current depression detection technologies are mostly completed through traditional methods such as patient interviews and PHQ scores [3], which are inefficient and have low accuracy. Later, researchers completed basic feature representation, static word embedding, and contextual word embedding through machine learning [4], and [5] used recurrent neural networks to identify depression from text, semantics, and written content. However, this unimodal model only targets text data, and performs multimodal processing on structured data, which makes it less comprehensive in depression detection and reduces the accuracy.

This paper proposes an RBF kernel BERT Transformer multimodal model. First, the text data and structured data are preprocessed respectively, and then feature extraction is performed through the text feature extraction module and the structured data extraction module respectively. The text data processed using the BERT Transformer model, TF-IDF vectorization model, and depression sentiment lexicon analysis model is then concatenated with the structured feature data after feature extraction to form a hybrid feature vector. Finally, the fused feature vector is fed into an SVM classifier to calculate the depression index and depression probability, with specific data representing the depression prediction results.

The multimodal feature fusion framework proposed in this paper innovates in model framework design, enabling multimodal processing of both text and structured feature data. This improves the comprehensiveness and data diversity of depression detection. Furthermore, the introduction of a weighted depression lexicon enhances the sensitivity of high-risk depression sentiment words, significantly improving detection accuracy. Finally, by quantifying and grading the depression index, targeted recommendations can be provided for groups with varying degrees of depression.

2. Methods

2.1. Overall Network Structure

This paper uses Bert Transformer as the semantic extraction method and RBF kernel SVM as the classifier. The overall process is shown in Figure 1.

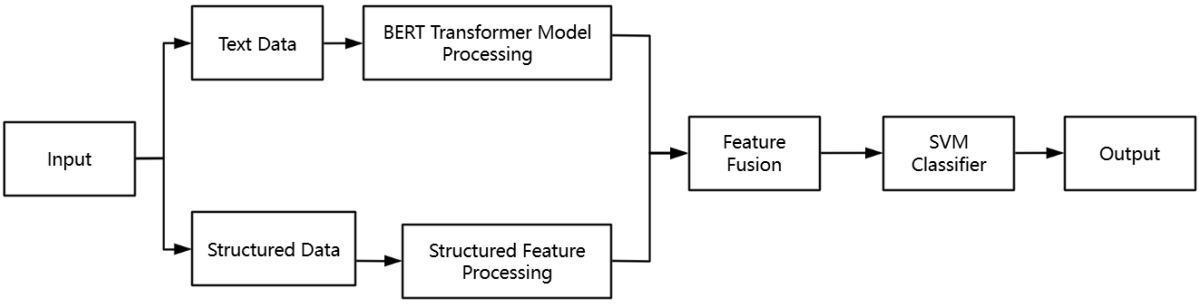

Fig. 1 Overall network structure diagram (picture credit: Original)

The model input consists of two core data types: user text data, including their feelings and mental health status, and representative structured feature data about the user (such as age, academic stress, and daily sleep duration). After preprocessing, the text and structured data are fed into the Bert Transformer text extraction module and structured feature extraction module, respectively, for feature extraction. The text and structured features are concatenated using np.concatenate to form a mixed feature vector. The mixed feature vector is then normalized using StandardScaler to eliminate the influence of different feature scales. Finally, the fused feature impact is fed into the SVM classifier to calculate the probability of depression and depression index.

2.2. Text Feature Extraction Module

After preprocessing, user text data will be simultaneously extracted through three models and finally fused, as shown in Figure 2:

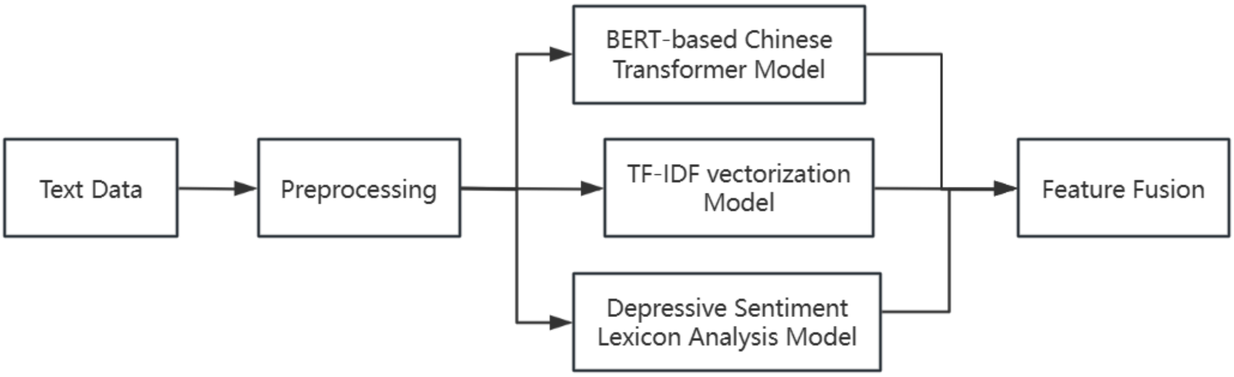

Fig. 2 Text feature extraction structure diagram (picture credit: Original)

The text data is first preprocessed: first, the text is denoised to remove invalid characters in the text. Then, the Chinese text is segmented using Jieba, and words and punctuation marks without emotional meaning are removed, retaining the core emotional words. Finally, the text is converted to string format and empty text is processed. For BERT input, the text length needs to be unified to 20 tokens through padding and truncation.

The BERT Transformer model is responsible for extracting in-depth semantic features of the text. It adopts an encoder architecture (Figure 3), where N is the number of encoders or decoders. [6] Compared with the original transformer model, this model abandons the decoder, output embedding, and final classification head, and adds segment embedding and special tokens, which enhances the ability to capture sentence-level semantics and text boundaries, making it more suitable for natural language understanding tasks (such as semantic analysis of depression text).

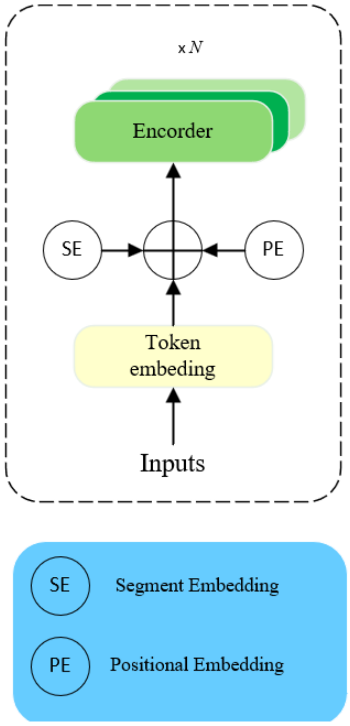

Fig. 3 BERT Transformer model structure diagram (picture credit: Original)

The TF-IDF vectorization model can filter and save sentiment-featured words and eliminate irrelevant words. The process is shown in Figure 4.

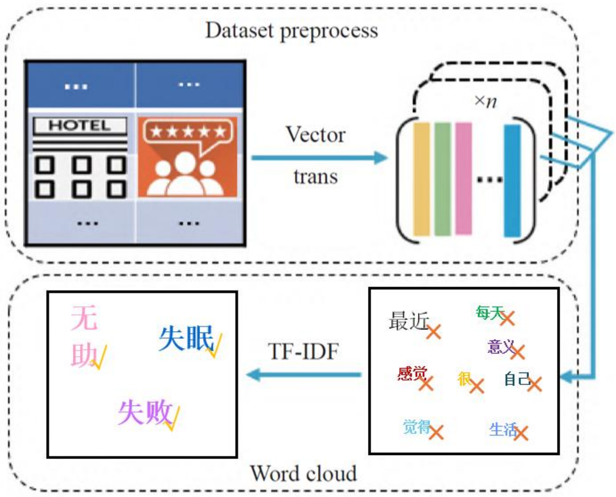

Fig. 4 TF-IDF vectorization processing structure diagram [6]

First, the vector obtained after preprocessing is fed into the TF-IDF vectorization model, where the following algorithm is used to filter sentiment words. The formula for Calculating word frequency (TF) is as follows.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k nk,j} \tag{1}$$

Where Term Frequency (TF) represents the frequency of a term (keyword) in the text, $n_{i,j}$ is the number of times the term appears in file d_j , and the denominator is the sum of the number of times all words appear in file d_j .Calculating reverse file frequency (IDF)'s formula is

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \tag{2}$$

The larger the IDF, the better the ability of the term to distinguish categories. Here, $|D|$ is the total number of documents in the corpus. $|\{j: t_i \in d_j\}|$ represents the number of documents containing the term t_i (i.e., the number of documents where ni, j ≠0) .Calculating TF-IDF is

$$TF - IDF = TF \times IDF \tag{3}$$

High word frequency within a specific document, combined with low word frequency across the entire document collection, can result in a high TF-IDF weight. Therefore, TF-IDF tends to filter out common words while retaining important ones.

The execution process of the sentiment dictionary is shown in Figure 5. First, text is input. Data preprocessing (including denoising and removing invalid characters) is performed, followed by word segmentation. Then, words of varying types and degrees from the depression sentiment dictionary are fed into the model for training. Finally, the sentiment type is output based on the sentiment judgment.

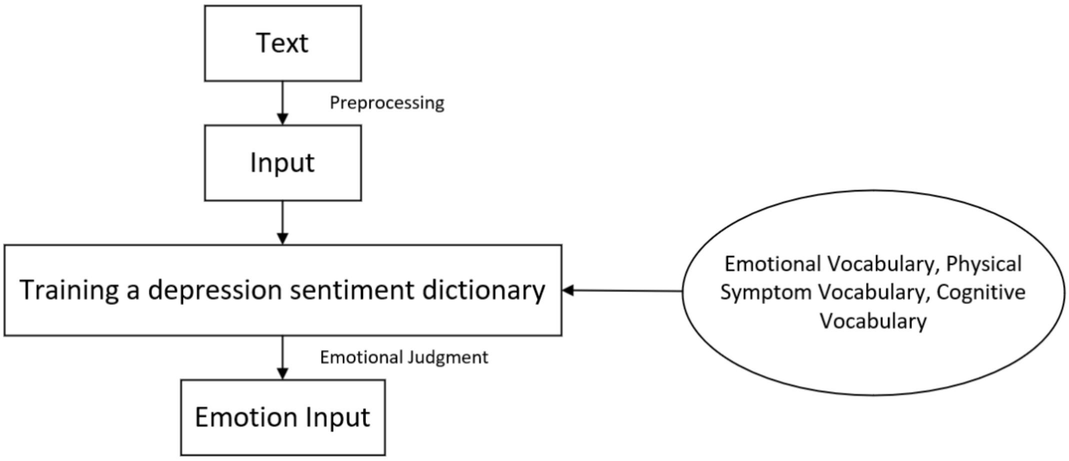

Fig. 5 Sentiment dictionary complex analysis model diagram [7]

2.3. Structured Feature Processing Module

The input structured feature data will be processed by the structured feature processing module to obtain a structured feature dictionary, which will then be converted into a 13-dimensional array for subsequent feature fusion with text features. The process of the structured feature processing module is shown in the figure 6:

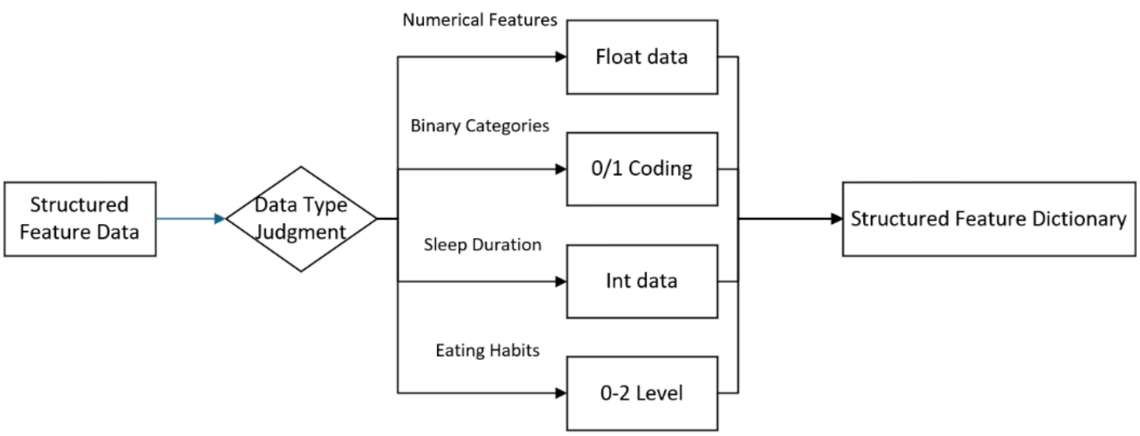

Fig. 6 Structured feature processing diagram (picture credit: Original)

During data preprocessing, structured feature data is first mapped to a mapping dictionary based on defined feature mapping rules. This data is then converted into data for subsequent data type determination. For example, "healthy" in dietary habits is mapped to 2, and "5 hours" in sleep duration is mapped to 5. The data type determination then categorizes each type into four types: float data, 0/1 encoding, int data, and feature data with a scale of 0-2. Finally, the structured data after feature extraction and processing is merged into a 13-dimensional array for subsequent processing by subsequent modules.

2.4. SVM Classification Module

The 969-dimensional text feature vector and the 13-dimensional structural feature vector are concatenated through np.concatenate to form a 982-dimensional mixed feature vector. Then, the mixed feature vector is standardized using StandardScaler (mean is 0, variance is 1) to eliminate the influence of different feature scales. [8] Finally, the constructed user vector is sent to the SVM classifier for processing.

After the weighing of text feature words, the depression score formula of SVM [9] is calculated as:

$$DepressionScore = \sum_{t \in T} wt \cdot \delta(t \in D) \tag{4}$$

Where T represents the set of terms after text segmentation, wt represents the sentiment weight of term t, and $\delta(t \in D)$ is an indicator function, where the function value is 1 when $t \in D$ and 0 otherwise. The following formula is used to calculate depression density:

$$DepressionDensity = \frac{DepressionScore}{max(|T|, 1)} \tag{5}$$

The proportion of depressive words in a text is measured by dividing the weighted depression score by the maximum word count T. [10] For example, for a simple text segmentation T = ["I", "feeling", "very", "painful", "and", "helpless"], where "painful" has a weight of 0.95 and "helpless" has a weight of 0.9, the depression score for this text is calculated to be DepressionScore (T) = 0.95 + 0.9 = 1.85, and the depression density is DepressionDensity (T) = 61.85 ≈0.308.

3. Experimental Analysis

3.1. Experimental Environment and Dataset

This experiment was run on a NVIDIA GeForce RTX 4060 laptop GPU and a 13th Gen Intel(R) Core (TM) i9-13950 CPU@2.2GHz, using Windows 11 64-bit. The model was built using Python version 3.9.23, trained with PyTorch version 2.5.1+cu121 and CUDA version 12.1, using Anaconda 24.11.3. TF-IDF and Transformer were used for feature extraction, and Support Vector Machine (SVM) was used for training.

This paper uses the self-developed student depression detection dataset, Depression Predictor. This dataset is structured as a CSV file and contains a series of features that record various indicators of depressed and non-depressed students, as well as Chinese text data. It contains 27,902 data points, each of which includes an ID: a unique identifier for each student. The dataset includes information such as age, gender, city, CGPA (grade point average), sleep duration, occupation, work stress, academic stress, learning satisfaction, job satisfaction, and dietary habits. The Depression column records the student's current depression status with "1" and "0." The User_Text column records the text language of the depressed individual. To fully train the model and prevent overfitting, this article uses the first 15,000 samples in the dataset for training and the last 10,000 for testing.

3.2. Test Indicator Settings

3.2.1. Depression indicators

This paper improves the original mSVM and performs TF-IDF and depression word weighting. Depression words come from a simple depression emotion dictionary (Table 1), in which each emotion feature word contains a different depression weight.

Table 1. Depression Lexicon

Category	Term	Weight	Category	Term	Weight
Affective	“悲伤”	0.90	Somatic	“头疼”	0.70
	“悲哀”	0.90		“头晕”	0.70
	“痛苦”	0.95		“失眠”	0.80
	“抑郁”	0.95		“发力”	0.75
	“郁闷”	0.85		“没食欲”	0.75
	“焦虑”	0.90		“心慌”	0.70
	“内疚”	0.80		“胸闷”	0.70
	“自责”	0.85		“疲惫”	0.75
	“绝望”	0.95		“嗜睡”	0.70
	“孤独”	0.85		“酸痛”	0.65
	“无助”	0.90		“恶心”	0.70
	“恐惧”	0.80		“呕吐”	0.70
	“压抑”	0.85		“发抖”	0.65
	“沮丧”	0.90		“冷汗”	0.65
	“崩溃”	0.95		“心悸”	0.75
	“心碎”	0.90	Cognitive	“无用”	0.85
	“失落”	0.85		“失败”	0.90
	“委屈”	0.80		“没价值”	0.90
	“自卑”	0.85		“不如死”	0.95
	“痛苦不堪”	0.95		“自杀”	1.00
				“绝望”	0.95
				“没意思”	0.80
				“空虚”	0.80
				“自责”	0.85
				“嫌弃”	0.80
				“相死”	0.97
				“活够了”	0.95
				“结束一切”	0.98
				“撑不下去”	0.92
				“毫无希望”	0.93

This article uses the depression index normalization:

$$DepressionIndex = \begin{cases} 20 \times DepressionDensity & d \leq 0.05 \\ 1 + 40 \times (d - 0.05) & 0.05 < d \leq 0.10 \\ 3 + 30 \times (d - 0.10) & 0.10 < d \leq 0.20 \\ 6 + 20 \times (d - 0.20) & 0.20 < d \leq 0.30 \\ 8 + min(20 \times (d - 0.30), 2) & d > 0.30 \end{cases} \quad (6)$$

A continuous, nonlinear, monotonically increasing function is used to segmentally amplify d and DepressionDensity. This maintains stability in the low-density depression range, amplifies small changes in the intermediate range to enhance sensitivity, and prevents saturation in the high-density range, capping the maximum score to facilitate stable modeling and threshold determination. The resulting DepressionIndex serves as an indicator of the subject's depression level. An Index < 3 indicates normal,3≤Index < 6 indicates a tendency toward depression, and an Index≥6 indicates confirmed depression.

3.2.2. Model indicators

According to the indicators required by the model, accuracy, recall rate, F1 score, and AUC formula are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

$$AUC = \int_{\mathbf{0}}^{\mathbf{1}} ROC(t)dt \tag{11}$$

Where: TP indicates that the true example is judged as a positive sample; FP indicates that the false example is judged as a positive sample; FN indicates that the false example is judged as a negative sample; TN indicates that the true example is judged as a negative sample. In the ROC formula, the horizontal axis is FPR and the vertical axis is TPR (true positive rate), which represents the classification performance under different thresholds.

3.3. Comparative Experiment

Test the trained model on an instance: {'Gender':'Female', 'Age':25, 'AcademicPressure':4, 'WorkPressure':0, 'CGPA':7.5, 'StudySatisfaction':3, 'Job Satisfaction': 0, 'Sleep Duration': '5-6 hours', 'Dietary Habits': 'Moderate','Have you ever had suicidal thoughts ?': 'Yes', 'Work/Study Hours': 8, 'Financial Stress': 3, 'Family History of Mental Illness': 'Yes', 'User_Text': "I feel very helpless recently. I suffer from insomnia every day. I feel like a failure and that my life has no meaning."} The result is a text depression index of 6.00/10.0 and a comprehensive depression level of: Severe depression, with relevant advice provided (Recommended reading therapy: 1. "Depression Self-Help Handbook" - Professional Guide 2. Crisis Intervention Hotline Information 3. Directory of Professional Psychological Counseling Resources 4. Emergency Response Plans. Recommendation: Seeking professional help is strongly recommended, in conjunction with reading therapy. Warning: High depression risk detected, please seek professional help immediately.

The RBF kernel BERT Transformer model described in this article was compared with the Transformer model and the TF-IDF + Transformer model, and their respective model metrics were tested. As shown in the AUC values for the training set, the RBF kernel BERT Transformer model shows a range of variation between 0.98 and 1.00, with most steps at 0.99 and 1.00 exceeding the other two models. The Transform model is in the range of 0.97-0.98, and the TF-IDF Transformer is mostly in the range of 0.98-0.99. This is because the 982-dimensional features of the RBF kernel BERT Transformer model provide richer discriminant information, making it easier to separate the prediction scores of positive samples from negative samples, thereby improving the ROC(t)dt value and making the parameters more accurate.

Table 2. Val AUC comparison table (divided by Epoch)

Epoch	Transformer	TF-IDF Transformer	RBF-TF-IDF Transformer
Training set	0.97	0.98	0.99
Test set	0.98	0.99	1.00

The accuracy, recall, and F1 score of the RBF kernel BERT Transformer model are similar to those of the other two models, fluctuating between 1 and 0.99. This is because the extraction logic and classifier parameters of the three models are consistent.

4. Conclusion

To address issues such as the original Transformer model's lack of multimodal feature fusion, low sensitivity to high-risk depressive sentiment words, and inability to quantify depression severity, this paper proposes a depression detection model based on an improved RBF kernel BERT Transformer. For text data, this paper uses the BERT model to extract deep semantic features from the text, a TF-IDF vectorization model to remove non-sentimental words, and a depression sentiment lexicon analysis model to output sentiment. After integrating structured data features, a user vector and a multi-dimensional regularized SVM are constructed. The weights of each word in the depression sentiment lexicon are used to calculate the probability and severity of depression.

The diversity and comprehensiveness of the structured and textual data in this paper's dataset remain to be improved, and the model can be subsequently distilled and scaled to reduce computational complexity. Experimental results show that the proposed RBF kernel BERT Transformer depression detection model achieves an AUC value ranging between 0.99 and 1, with precision, recall, and F1 scores all close to 1, outperforming most current single-modal depression detection models. This provides highly accurate technical support for the detection of depression in adolescents and further responds to society's needs for the prevention and treatment of adolescent depression.

References

[1] Cassano P, Fava M. Depression and public health: an overview. Journal of Psychosomatic Research, 2002, 53 (4): 849-857.

[2] Klein D N, Goldstein B L, Finsaas M. Depressive disorders. Child and Adolescent Psychopathology, Third Edition, 2017: 610-641.

[3] Marriwala N, Chaudhary D. A hybrid model for depression detection using deep learning. Measurement: Sensors, 2023, 25: 100587.

[4] Xu Dongdong, Cai Xiaohong, Liu Jing, et al. A review of depression detection research based on social media text data. Journal of Computer Engineering & Applications, 2023, 59 (4).

[5] Amanat A, Rizwan M, Javed A R, et al. Deep learning for depression detection from textual data. Electronics, 2022, 11 (5): 676.

[6] Gao Jiaxi, Huang Haiyan. Text sentiment analysis based on TF-IDF and multi-head attention Transformer model. Journal of East China University of Science and Technology (Natural Science Edition), 2024, 50 (01): 129-136.

[7] Wang Yingjie, Zhu Jiuqi, Wang Zumin, et al. A review of natural language processing applications in text sentiment analysis. Journal of Computer Applications, 2022, 42 (04): 1011-1020.

[8] Han Meihua, Zhao Jingxiu. Research on reading therapy model based on “user profile”: A case study of depression. Journal of Academic Libraries, 2017, 35 (06): 105-110.

[9] Wang Yao, Jia Baolong, Du Yining, et al. Multi-dimensional regularized SVM-based depression tendency detection method on social networks. Computer Applications and Software, 2022, 39 (03): 116-120.

[10] Yang Junzhe, Song Ying, Chen Yifei. Text sentiment analysis model integrating topic features. Computer Science, 2024, 51 (S1): 171-178.