Multi-Dimensional Digital Media Sentiment Visualization Intelligent Analysis System Based on Machine Learning Algorithm

Mengwei Lei[1]*, Qiong Chen[2]
School of Visual Art, Hunan Mass Media Vocational and Technical College, Changsha, Hunan 410100, China[1]
School of Materials Science and Engineering, Central South University, Changsha, Hunan 410083, China[2]

***Abstract*—This study builds a multi-dimensional sentiment analysis system to solve the problem of sentiment prediction of text and image data in the Weibo platform. By combining CNN (Convolutional Neural Network), BiLSTM (Bidirectional Long Short-Term Memory) and Attention mechanism (AM), the accuracy of sentiment classification is improved, which helps to better understand and analyze user sentiment expressions in social media. This study uses crawler tools to collect text and image data of 1,000 users on the Weibo platform from January to December 2021 to ensure the diversity and representativeness of the data; the text data is segmented, stop words are removed, and the text is converted into vectors; at the same time, the ResNet-50 pre-trained model is used to extract the deep features of the image, CNN is used to process the image data, and BiLSTM captures the contextual information in the text data. Finally, the AM is used to enhance the model's attention to emotional expression. Experimental results show that the proposed Word2Vec (Word to Vector) model performs outstandingly in the accuracy of sentiment classification. The accuracy of the CNN-BiLSTM-Attention model in positive, neutral and negative classification tasks is 97.5 per cent, 95.4 per cent and 91.6 per cent, respectively, which are significantly better than the performance of the CNN and BiLSTM models, especially in the evaluation indicators such as accuracy and macro F1. This study proposes a multimodal sentiment analysis system based on CNN-BiLSTM-Attention, which significantly and effectively improves the accuracy of social media sentiment classification. The system can effectively process complex sentiment categories and multimodal data, and has broad application prospects, especially in the fields of social media sentiment analysis and public opinion monitoring.**

***Keywords*—Digital media; sentiment analysis; intelligent systems; multimodal data**

I. Introduction

Against the backdrop of the rapid development of digitalization and intelligence [1], social media [2] has become an important carrier for users to express their emotions and disseminate information. The multimodal data contained in social media provides rich resources and challenges for sentiment analysis research. Weibo [3] is one of the largest social media platforms in China. Its user-generated content includes text, images and other data forms. These data not only reflect the users' emotional states and psychological activities, but also reveal the dynamic changes in social public opinion. However, traditional sentiment analysis methods are usually limited to single-modal data processing, such as analyzing only text or image features, and cannot fully utilize the semantic complementarity and correlation between multimodal data. Weibo data is unstructured, with diverse and metaphorical emotional expressions, which increases the complexity and difficulty of sentiment classification. To address the above problems, this study constructs a multi-dimensional sentiment analysis system based on the CNN-BiLSTM-Attention model. CNN [4] is used to extract image features, BiLSTM [5] is used to capture the time series features of text, and the AM [6] is used to focus on key sentiment information, thus achieving high-precision sentiment classification of Weibo multimodal data. This research has broad application value in the fields of business intelligence decision-making, personalized recommendation systems, and social and psychological health monitoring, and provides an important theoretical basis and practical reference for the design and implementation of future intelligent sentiment analysis systems.

This study innovatively integrates the data processing of two modalities, image and text, and constructs a multi-dimensional sentiment analysis system by combining CNN, BiLSTM and AM to solve the key problems in the sentiment classification of multimodal data on the Weibo platform. In view of the complexity of multimodal data and the diversity of emotional expression, this study introduces an analysis framework that can efficiently integrate multimodal features, providing a new solution for sentiment classification; by optimizing the model structure and introducing advanced deep learning mechanisms, especially the effective application of the AM, the model's ability to focus on key emotional features is enhanced, thereby significantly improving the accuracy and robustness of sentiment classification. This study shows innovation in multimodal data processing, fully exploring and integrating the semantic complementarity and correlation between different modalities, and providing technical support for the modeling of complex emotional expressions. On the theoretical level, this study expands the research framework of multimodal sentiment analysis and deepens the understanding of sentiment feature extraction and classification methods; on the practical level, the proposed method has strong adaptability and scalability, and provides a practical solution for social media sentiment monitoring, public opinion analysis, and intelligent decision support. By optimizing and innovating existing methods, the data processing of image and text modes is innovatively integrated, which not only provides a theoretical basis for the development of sentiment analysis technology, but also opens up a new direction for research and application in related fields, which has important academic value and practical significance.

II. Related Work

The research on multi-dimensional digital media emotion visualization [7] intelligent analysis system is a cutting-edge direction in the intersection of artificial intelligence and big data analysis. It aims to achieve comprehensive mining and intuitive presentation of emotional information in digital media by integrating multimodal data [8] processing technology and deep learning algorithms [9]. In recent years, sentiment analysis has expanded from traditional unimodal text analysis to multimodal data processing, and the research focus has gradually shifted to how to efficiently extract and fuse multimodal features such as text and images to improve the accuracy and robustness of sentiment classification. In this context, CNN has become the core technical support for multimodal sentiment analysis systems due to its excellent performance in image feature extraction. BiLSTM has become the core technical support for multimodal sentiment analysis systems due to its ability to accurately capture text time series features, and AM has become the core technical support for multimodal sentiment analysis systems due to its advantage in focusing on key features. As an important part of research, sentiment visualization greatly improves the interpretability and user experience of sentiment analysis results by graphically presenting the distribution, trend and structure of sentiment data. Existing research has shown wide application value in fields such as social media sentiment monitoring, public opinion analysis and business intelligence decision-making. However, the semantic differences between multimodal data and the complexity of feature fusion are still difficulties in current research. Therefore, how to build an efficient, accurate and scalable multi-dimensional sentiment analysis system has become the key to promoting the development of this field.

The application of machine learning algorithms in multi-dimensional sentiment analysis systems is an important research direction in the current field of artificial intelligence and big data analysis [10], dedicated to solving the complexity of sentiment data and multi-modal feature fusion problems in digital media [11]. Digital media sentiment data usually exists in various forms such as text and images, and is characterized by high dimensionality, nonlinearity, and heterogeneity, which poses a severe challenge to traditional sentiment analysis methods. In text data processing, deep learning models such as BiLSTM can capture the contextual semantic relationship of text and dynamically focus on key sentiment-related content through the AM, thereby improving the accuracy of sentiment classification. In terms of image analysis, CNN effectively extracts emotion-related visual features through multi-layer convolution operations, especially in facial expression and scene emotion analysis. The key to a multimodal emotion analysis system lies in feature fusion. Machine learning achieves deep alignment and fusion at the semantic level through joint modeling of text and images, and enhances the ability to understand complex emotional expressions. Sentiment visualization technology combines dimensionality reduction and clustering algorithms [12] to transform high-dimensional sentiment data into intuitive graphical representations, providing users with clear insights into sentiment distribution and trends. Advanced deep learning methodologies such as LSTM, CNN, and BERT have been explored to enhance sentiment interpretation within natural language processing, as demonstrated by Aniket Kulkarni et al. (2024). Our framework incorporates their transformer-based advancements along with contextual embedding techniques to evaluate sentiment across multiple digital formats, thereby improving accuracy, scalability, and offering richer, multi-dimensional visualization capabilities [13]. A sentiment classification system combining Levy distribution-based Dung Beetle Optimization (LDBO) with Support Vector Machine (SVM) focuses on social media content, following the approach introduced by Layth Hussein et al. (2024). By leveraging this optimization technique, our model enhances feature extraction and hyperparameter tuning for multi-modal sentiment analysis, resulting in superior classification performance and operational efficiency that supports real-time, intelligent sentiment representation [14]. Mohan Reddy Sareddy (2023) investigated cloud-based CRM infrastructure strategies aimed at fostering business success within digital ecosystems. Building on this foundation, our research integrates machine learning models for comprehensive sentiment assessment and multi-format visualization across various digital channels, which enables rapid and actionable insights that support strategic decision-making and enhance engagement beyond traditional CRM functionalities [15]. Peng Yang (2019) developed a phishing detection framework employing deep learning techniques such as multidimensional feature extraction and sequential analysis. Adapting their BiLSTM and attention-based mechanisms, our system captures temporal sentiment dynamics within intricate digital media datasets, emphasizing emotionally significant features, which leads to enhanced classification accuracy and more nuanced, interpretable sentiment visualizations [16]. An ensemble blending method to advance behavioral data analysis through the combination of multiple machine learning models was proposed by Akhil Raj Gaius Yallamelli et al. (2025). Our platform utilizes this ensemble blending strategy to enhance sentiment analysis across varied digital media, thereby increasing prediction accuracy and delivering more detailed, insightful visualizations of user sentiment [17]. The comprehensive application of these technologies not only promotes the technological progress of multi-dimensional sentiment analysis systems but also provides solid technical support and practical significance for social media public opinion monitoring, brand sentiment analysis, and user behavior prediction.

III. Methods

A. Data Source

The data in this study comes from the Weibo platform. The reason for choosing this platform is that its users are highly active, the content is diverse, and it can provide rich emotional expression samples. The specific data collection range is from January to December 2021, ensuring that the data covers a full annual cycle to capture the trend of sentiment changes in different time periods. The data volume comes from the Weibo content of 1,000 users, who are selected based on their activity level and content diversity to improve the representativeness of the data and the emotional coverage. The data types collected include text and images, which are used to analyze emotional expressions at the language and visual levels respectively.

In order to efficiently obtain Weibo data, advanced web crawler technology [18-19] was used to collect Weibo content from 1,000 users, following the platform's terms of service and data collection specifications to ensure the legality and compliance of data collection. The crawler tool simulates user behavior and regularly captures the target user's Weibo content, including text, images, comments, forwarding numbers, and likes, to construct a multi-dimensional sentiment analysis dataset. In the user screening stage, user groups with different genders, ages and regional distributions were selected to ensure the diversity and representativeness of the sample. In addition, through keyword filtering and emotional label pre-classification, the collected data was guaranteed to cover microblog content with positive emotions, negative emotions and neutral emotions. The collected text data is shown in Table I.

TABLE I. TEXT DATA DISPLAY

User	Text	Emotional tendency	Category
1	I am so happy to have found a satisfying job!	Positive	Joy
2	Thank you friends for your continued support, I am deeply touched.	Positive	Gratitude
3	I won first place in the campus sports meet today and I feel very proud.	Positive	Pride
4	I didn't react much when I heard the news, I just thought it was quite normal.	Neutral	Calm
5	This movie was kind of boring and started to distract.	Neutral	Boring
6	A bit confused as to why this question is so complicated.	Neutral	Puzzled
7	I'm so fed up with the delays on this project, it's so infuriating!	Negative	Anger
8	After hearing the news, I felt very sad and heavy in my heart.	Negative	Sad
…	…	…	…
1000	Life has been a mess lately and everything feels wrong.	Negative	Anxiety

B. Data Preprocessing

Text data preprocessing includes word segmentation, stop word removal, sentiment vocabulary tagging and text vectorization. Weibo texts are mostly unstructured data and need to be segmented to split continuous Chinese character sequences into word sequences. The Chinese word segmentation tool Jieba is used, which is based on the hidden Markov model and achieves efficient word segmentation through probability statistics and dictionary matching. The word segmentation formula is expressed as:

$$P(W|S) = \prod_{i=1}^{n} P(w_i|s_i) \tag{1}$$

where, W represents a word sequence, S represents a state sequence, w_i and s_i represent a word and a state respectively.

Stop words refer to words that have no practical meaning in semantic analysis. By matching the stop word list, these words are filtered out from the word segmentation results, and words with emotional meaning are retained. Sentiment word tagging refers to tagging sentiment words in texts using a method based on sentiment dictionaries. Sentiment dictionaries include positive words, negative words, and degree adverbs. The formula for calculating sentiment intensity is expressed as:

$$S = \sum_{i=1}^{n} (P_i \cdot M_i) \tag{2}$$

P_i represents the polarity weight of the ith sentiment word, and M_i is the weight correction value of the degree adverb.

Text vectorization is to convert text into a vector form using the Word2Vec model. Word2Vec trains word vectors through the skip-word model, which can capture the semantic relationship between words. The objective function is expressed as:

$$J = -\frac{1}{T} \sum_{t=1}^{T} \sum_{-c \le j \le c, j \ne 0} \log P(w_{t+j}|w_t) \tag{3}$$

T represents the corpus length, c is the context window size, w_t and w_{t+j} are the center word and context word, respectively.

Image data preprocessing includes image preprocessing and feature extraction. To ensure the consistency of model input, all images are adjusted to 224×224 pixels to adapt to the input requirements of the deep learning model, and median filtering is used to remove noise in the image and enhance image clarity. The median filtering formula is expressed as:

$$f(x, y) = \text{median}\{g(i, j)\}, (i, j) \in N(x, y) \tag{4}$$

where, $g(i, j)$ represents the pixel value in the area $N(x, y)$, and $f(x, y)$ is the pixel value after filtering. In order to accelerate model training and improve convergence speed, normalization is also required to scale the pixel value to the range of [0, 1]. The normalization formula is expressed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

Feature extraction refers to extracting deep features of an image using a pre-trained ResNet-50 model [20]. Through the above method, both text and image data are converted into a unified vector form, laying a solid foundation for the subsequent training of sentiment analysis models and multimodal fusion.

C. Model Design and Construction

In this study, CNN is used to process image data in Weibo to help extract emotion-related visual features. CNN contains convolutional layers and pooling layers. The core idea of the convolution operation is to use multiple filters to slide the image, calculate the weighted sum of the local area, and generate a feature map. The convolution operation formula is as follows:

$$F(x, y) = (I * K)(x, y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} I(x + m, y + n) K(m, n) \tag{6}$$

The formula for maximum pooling is as follows:

$$P(x, y) = \max_{m,n \in R} F(x + m, y + n) \tag{7}$$

where, R represents the size of the pooling window, and $P(x, y)$ is the feature map after pooling. Through multi-layer convolution and pooling operations, CNN can gradually extract low-level to high-level spatial features in the image and significantly reduce the number of parameters, thereby improving computational efficiency.

BiLSTM can learn the contextual information of the text from both forward and reverse directions, so as to more comprehensively understand the emotional expression of the text. LSTM units have memory functions and can maintain and update important information in long time series. The calculation formulas of LSTM are expressed as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{9}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{10}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{12}$$

$$h_t = o_t * \tanh(C_t) \tag{13}$$

The AM aims to improve the model's attention to key parts by calculating the weights of each part of the input sequence. When processing long sequence inputs, it can avoid information loss and improve the accuracy of sentiment analysis. The common attention calculation method is based on the additive model, in which, given the query vector q and the key vector k, the attention weight is calculated as:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{14}$$

Among them, e_i=score (q, k_i) is the similarity score between the query vector and the key vector. The commonly used scoring function is expressed as:

$$e_i = w^T \tanh(W_q q + W_k k_i) \tag{15}$$

where, W_q and W_k represent weight matrices, and w is a trainable parameter. After obtaining the weights, the attention output is obtained by weighted summation:

$$\text{Attention Output} = \sum_{i=1}^{n} \alpha_i \cdot v_i \tag{16}$$

where, α_i represents the value vector, and v_i is the corresponding weight. The AM enables the model to dynamically adjust its attention to the input according to the intensity of the emotion, which can effectively improve the performance of the model for long text or complex image data.

D. Multi-Dimensional Sentiment Analysis System

The multidimensional sentiment analysis system is an intelligent system that comprehensively utilizes advanced technologies such as natural language processing, computer vision, deep learning, and machine learning to process and integrate multimodal information of different data types to perform sentiment analysis on digital media content such as social platforms. Through the combination of deep learning, machine learning and other advanced technologies, the system can fully understand and predict user emotions from a multi-dimensional perspective and provide valuable emotional information for decision makers. The system interface constructed in this study is shown in Fig. 1.

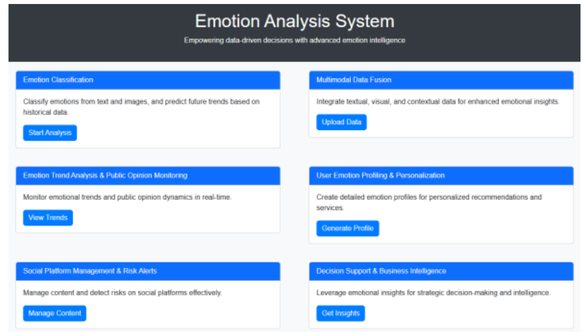

Fig. 1. System interface

The multi-dimensional digital media emotion visualization intelligent analysis system in this study combines advanced machine learning and deep learning technologies to perform efficient emotion analysis and prediction on multimodal data on Weibo, thereby achieving accurate emotion expression recognition and emotion trend visualization. The analysis system uses advanced technologies such as natural language processing, computer vision and deep learning, and has many innovative functions, including sentiment classification and prediction, multimodal data fusion, sentiment trend analysis and public opinion monitoring. Its core role is to help all types of users make data-driven decisions in the rapidly changing social media environment through accurate sentiment recognition and trend monitoring, improve operational efficiency, enhance user experience, optimize social public opinion management, and provide strong emotional intelligence support for various industries.

E. Model Training and Optimization

The distribution of the data set is shown in Table II.

TABLE II. DATA SET DISTRIBUTION

Category	Training set	Validation set	Test set
Joy	714	204	102
Gratitude	686	196	98
Pride	854	244	122
Calm	763	218	109
Boring	707	202	101
Puzzled	609	174	87
Anger	651	186	93
Sad	658	188	94
Anxiety	672	192	96

Emotional categories are classified into three tendencies: positive, neutral, and negative. The data composition is shown in Table III.

TABLE III. EMOTIONAL TENDENCY DATA

Tendency	Training set	Validation set	Test set
Positive	2254	644	322
Neutral	2079	594	297
Negative	1981	566	283

The initialization parameters for model training in this study are shown in Table IV.

TABLE IV. MODEL INITIALIZATION PARAMETERS

Initialization content	Effect	Value
Convolution kernel size	Extract local features of images	3×3
Number of convolutional layers	Depth of feature extraction	4
Convolution stride	Control the sliding step size of the convolution kernel	1
Pooling window size	Reduce feature map dimension	2×2
Number of LSTM hidden units	Controlling the memory capacity of BiLSTM	128
Number of LSTM layers	Enhance the model's ability to learn long sequence dependencies	2
Number of Attention Heads	Improve the model's ability to focus on key emotional information	8
Dropout probability	Preventing Overfitting	0.5
Learning Rate	Control the step size of model parameter updates	0.001
Optimizer	Efficiently optimize model parameters	Adam

The sentiment analysis model in this study mainly uses Dropout and L2 regularization. Dropout is a technique that randomly discards neural network nodes during training, with the aim of reducing the dependency between neurons and preventing the model from overfitting. The specific approach is to randomly discard the output of a part of the neurons during each training.

Accuracy is one of the most commonly used classification indicators, indicating the proportion of samples predicted correctly by the model of the total samples. The macro F1 value can effectively deal with the problem of class imbalance by calculating the F1 value of each category and then finding its arithmetic mean. The formula is:

$$\text{Macro F1}=\frac{1}{C}\sum_{i=1}^{C}\text{F1}_i \tag{17}$$

C represents the total number of categories. By introducing regularization technology and Adam, the model in this study can effectively avoid overfitting and accelerate convergence in sentiment analysis tasks, improving the accuracy and stability of sentiment classification. At the same time, accuracy and macro F1 value as evaluation indicators can fully reflect the performance of the model. When facing multi-category sentiment classification tasks, the macro F1 value can more fairly evaluate the overall performance of the model.

IV. RESULTS AND DISCUSSION

A. Sentiment Classification Performance

In order to evaluate the sentiment classification effect of different models when processing multimodal data on the Weibo platform, the accuracy is compared. The comparison results of the sentiment classification accuracy of different models are shown in Fig. 2.

According to the provided sentiment classification accuracy data, it can be clearly seen that the performance of the model has been significantly improved with the improvement of the architecture, and the CNN-BiLSTM-Attention model performs significantly better than other models in the sentiment classification task, with accuracies of 97.5 per cent, 95.4 per cent and 91.6 per cent for positive, neutral and negative sentiments respectively. This significant improvement is mainly due to the introduction of the AM. CNN extracts local features of images through convolutional layers. Although it can process image information, it is not capable of processing text data in sentiment analysis tasks, resulting in relatively low accuracy in the classification of negative and neutral sentiments. BiLSTM effectively captures contextual information in text through a bidirectional structure and can solve the problem of long-term dependency, but it is still limited in distinguishing complex emotions. The model can pay more attention to the parts with strong emotional expressions, thereby improving the accuracy of classification. The role of the AM is to avoid information loss in long texts or complex emotional expressions in traditional models through a weighted mechanism, allowing the model to more accurately capture subtle differences in emotions. In addition, reinforcement learning is performed on keywords, phrases, or contexts in each emotional category, which effectively improves the classification accuracy of complex emotions, thereby significantly improving the classification accuracy of positive and neutral emotions.

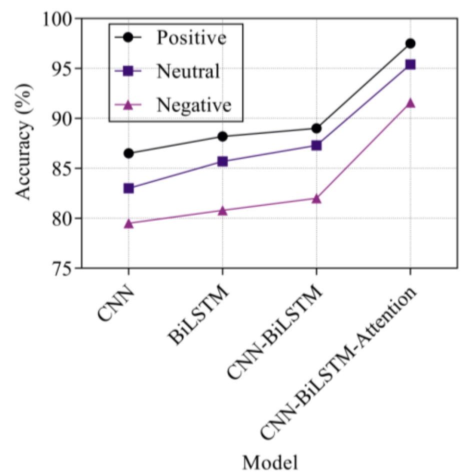

Fig. 2. Sentiment classification performance

B. Positive Sentiment Classification Ability

The comparison results of positive sentiment tendency classification macro F1 of different models are shown in Fig. 3.

From the data in Fig. 3, the macro F1 scores of the CNN-BiLSTM-Attention model in the three positive emotion categories of joy, gratitude and pride are 0.97, 0.94 and 0.91, respectively, which are significantly better than other models, mainly due to its effective focus on key emotional information and accurate capture of emotional expression in long texts. Although the CNN model can extract local features of images through convolutional layers, it has limited ability to process long-term dependencies and complex emotional expressions in texts, so its macro F1 score in sentiment classification is low, especially in pride, with a macro F1 of 0.74. BiLSTM captures the contextual information of texts through a bidirectional structure, which can better solve the limitations of traditional LSTM in processing long texts. In the classification tasks of joy, gratitude, and pride, the macro F1 scores have improved to 0.85, 0.83, and 0.79, respectively; however, BiLSTM still faces the problem of integrating image information and text information when processing multimodal data, so it has not achieved the best results in combining image and text information. After the CNN-BiLSTM-Attention model introduces the AM, it enhances the model's attention to key information by assigning different weights to different input parts. In situations where emotional expression is strong, it can effectively reduce the problem of information loss and improve the accuracy of sentiment classification. This shows that the AM not only improves the accuracy of the classification model but also improves the ability to extract emotional information in multimodal data fusion, so the CNN-BiLSTM-Attention model has a better macro F1 score for positive emotions than other models.

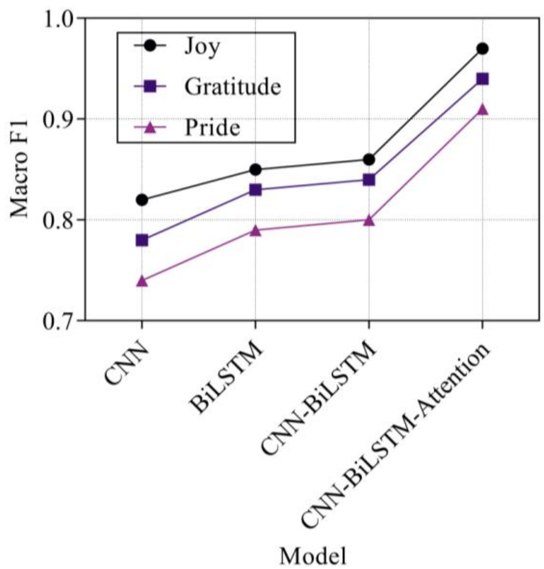

Fig. 3. Positive sentiment classification performance

C. Neutral Emotion Classification Ability

The comparison results of the neutral emotion tendency classification macro F1 of different models are shown in Fig. 4.

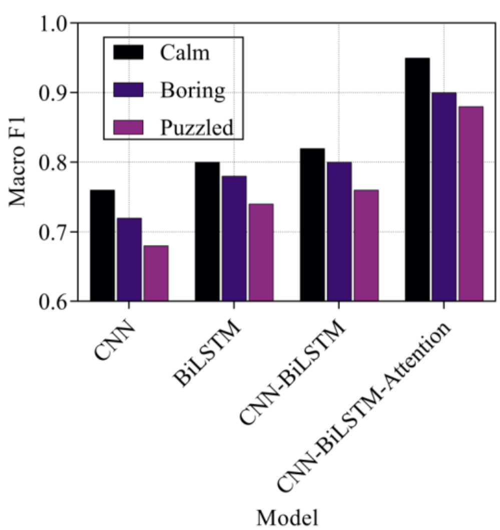

Fig. 4. Neutral sentiment classification performance.

As the model complexity increases, the macro F1 scores of each neutral sentiment category gradually improve; among them, CNN-BiLSTM-Attention performs best, with macro F1 of 0.95, 0.90, and 0.88 for calm, boring, and puzzled, respectively. The relatively low macro F1 score of the CNN model reflects that CNN has limited ability to understand emotions when processing text data, especially when processing complex neutral emotions, it cannot effectively extract contextual information, resulting in low classification accuracy. The macro F1 score of the BiLSTM model has been improved. Compared with CNN, BiLSTM captures the contextual information in the text through the bidirectional LSTM structure, which can better understand and analyze the subtle differences in neutral emotions, and has a stronger ability to capture long texts and complex emotions, but there is still room for improvement. The macro F1 score of the CNN-BiLSTM model has been further improved, especially in the performance of "calm" and "boring" emotions, which have been significantly improved to 0.82 and 0.80 respectively; after introducing the fusion of image and text information, the model can more comprehensively understand multi-dimensional emotional expressions, so its performance has been improved. With the introduction of the mechanism, the model can more accurately identify and classify neutral emotions with vague and complex emotional expressions, significantly improving the classification performance.

D. Negative Emotion Classification Ability

The comparison results of the negative emotion tendency classification macro F1 of different models are shown in Fig. 5.

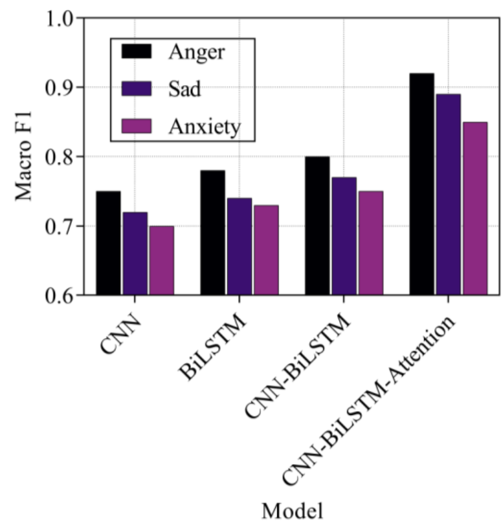

Fig. 5. Negative sentiment classification performance.

The macro F1 of CNN-BiLSTM-Attention for anger, sad, and anxiety are 0.92, 0.89, and 0.85, respectively, which reflects the enhanced feature extraction and information processing capabilities at a deeper level. The macro F1 score of the CNN model is relatively low, mainly because the convolution operation of CNN has limitations in the application of text data. The BiLSTM model has improved its performance in negative sentiment classification, and its score is higher than that of CNN. However, when processing large-scale sentiment data, it still faces the problems of uneven distribution of sentiment information and vague sentiment expression, which leads to insufficient recognition accuracy of the model in some negative sentiments. The CNN-BiLSTM model further combines the local feature extraction of CNN and the long-term dependency modeling of BiLSTM. Through multi-layer feature fusion, the model can overcome the limitations of CNN single convolutional layer feature extraction to a certain extent and improve classification accuracy. The most notable is the performance of the CNN-BiLSTM-Attention model, which has significantly improved its score on Macro F1. The AM significantly enhances the model's ability to capture key information by focusing on important information in the input data. In the recognition of negative emotions, the model can more accurately locate the key expression of emotions, effectively solving the problem of emotional information loss in long sequences, and ultimately improving the accuracy and robustness of classification. This improvement reflects the huge potential of multimodal information fusion and deep learning technology in sentiment analysis, and its superiority in complex sentiment analysis tasks.

V. Conclusion

This study successfully improved the sentiment classification performance of text and image data on the Weibo platform by building a multi-dimensional sentiment visualization intelligent analysis system based on the CNN-BiLSTM-Attention model, and achieved remarkable results in the recognition of negative, positive and neutral sentiment. By effectively combining a convolutional CNN, BiLSTM and Attention, this not only improves the accuracy and macro F1 score of sentiment classification, but also breaks through the limitations of traditional sentiment analysis methods in processing long sequence data, complex emotional expressions and multimodal information fusion. The main contribution of the study is to build an innovative multimodal sentiment analysis framework that can simultaneously utilize the features of text and image data, solve the problems of information diversity and context dependence in sentiment analysis, and optimize the model's sensitivity to sentiment intensity and subtle differences through the AM. This system not only provides a more accurate and efficient solution for sentiment analysis in social media, but also provides a new research perspective and technical path for fields such as sentiment computing, sentiment visualization and social network analysis. This study can help all kinds of enterprises and organizations to more accurately grasp user sentiment, optimize business decisions such as customer service and marketing, and promote the development of applications such as social opinion analysis and mental health intervention. However, this study also has some limitations, mainly in terms of the size of the dataset and the diversity of sentiment categories. The model is unstable when dealing with sentiment classification tasks on other platforms or in different language environments. Although the CNN-BiLSTM-Attention model performed well in this study, its computational complexity is high, and there is still room for optimization in training and inference time. Future research can further improve the generalization and real-time performance of the model by expanding the dataset size, enhancing cross-platform adaptability, and optimizing the model structure and algorithm efficiency, thus providing more efficient technical support for large-scale sentiment analysis applications.

Funding: Authors did not receive any funding.

Conflicts of interests: Authors do not have any conflicts.

Data Availability Statement: No datasets were generated or analyzed during the current study.

Authors' Contributions: Mengwei Lei, is responsible for designing the framework, analyzing the performance, validating the results, and writing the study. Qiong Chen, is responsible for collecting the information required for the framework, provision of software, critical review, and administering the process.

References

[1] L. Xiangjun, L. Xiaoyu, H. Xuebing, Y. Jiatao, and L. Rui, “Digital and intelligent technology of electrochemical energy storage power station and its application prospects,” Power Supply and Use, vol. 40, no. 8, pp. 3–12, 2023.

[2] [2] Y. Mengli, S. Wenhan, and Z. Bowen, “Research on the motivation of social media users' information deletion behavior - Taking WeChat Moments as an example,” J. Inf. Res. Manag., vol. 13, no. 4, pp. 84–95, 2023.

[3] C. Xingshu, C. Tianyou, W. Haizhou, Z. Zhilong, and Z. Jie, “Spatial-temporal analysis of the evolution of public opinion on the 'new crown pneumonia epidemic' based on Weibo data,” J. Sichuan Univ. (Nat. Sci. Ed.), vol. 57, no. 2, pp. 409–416, 2023.

[4] T. Yiyong and B. Chunyang, “Comment sentiment classification based on ALBERT-BiLSTM-CNN based on attention mechanism,” Modern Comput., vol. 30, no. 5, pp. 44–49, 2024.

[5] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, “A CNN-BiLSTM model with attention mechanism for earthquake prediction,” J. Supercomput., vol. 79, no. 17, pp. 19194–19226, 2023

[6] G. Yao, X. Haijun, Z. Zipeng, and L. Lu, “CNN-BiLSTM transient electromagnetic real-time inversion method based on attention mechanism,” Coalfield Geol. Explor., vol. 51, no. 10, pp. 134–143, 2023.

[7] [7] D. Jinhua, Z. Binrui, and Q. Xiaosong, “A review of research on artificial intelligence empowering cultural heritage field-Visual analysis based on CiteSpace,” Packag. Eng. Art Ed., vol. 44, no. 14, pp. 1–20, 2023.

[8] S. Weihao, Z. Yanfei, W. Junjue, Z. Zhuo, and M. Ailong, “Construction and application of flood disaster knowledge graph based on multimodal data,” J. Wuhan Univ. (Inf. Sci. Ed.), vol. 48, no. 12, pp. 2009–2018, 2023.

[9] Z. Yu, K. Wang, Z. Wan, S. Xie, and Z. Lv, “Popular deep learning algorithms for disease prediction: a review,” Cluster Comput., vol. 26, no. 2, pp. 1231–1251, 2023.

[10] Z. Xiaofan, H. Minghui, and X. Lei, “Analysis of energy management strategy of plug-in hybrid electric vehicles based on real vehicle test big data analysis,” J. Chongqing Univ., vol. 46, no. 2, pp. 11–29, 2023.

[11] T. Jie and T. Mengjie, “Research on the interactivity of virtual engine technology in digital media art teaching,” Educ. Theory Appl., vol. 6, no. 5, pp. 157–159, 2024.

[12] Z. Xuefeng, X. Qiang, T. Yanting, L. Jiayi, J. Jing, and Z. Zhiqiang, “Cloud architecture data center network abnormal traffic filtering algorithm based on improved grey clustering algorithm,” Telecommun. Sci., vol. 39, no. 7, pp. 90–98, 2023.

[13] A. Kulkarni, V. S. B. H. Gollavilli, Z. Alsalami, M. K. Bhatia, S. Jovanovska, and M. N. Absur, “Leveraging Deep Learning for Improved Sentiment Analysis in Natural Language Processing,” in 2024 3rd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), Nov. 2024, pp. 1–6, IEEE.

[14] L. Hussein, J. N. Kalshetty, V. S. B. Harish, P. Alagarsundaram, and M. Soni, “Levy distribution-based Dung Beetle Optimization with Support Vector Machine for Sentiment Analysis of Social Media,” in 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Aug. 2024, pp. 1–5, IEEE.

[15] M. R. Sareddy, “Cloud-Based Customer Relationship Management: Driving Business Success in the E-Business Environment,” International Journal of Marketing Management, vol. 11, no. 2, pp. 58–72, 2023.

[16] P. Yang, G. Zhao, and P. Zeng, “Phishing website detection based on multidimensional features driven by deep learning,” IEEE Access, vol. 7, pp. 15196–15209, 2019.

[17] A. R. G. Yallamelli, V. Mamidala, R. K. M. K. Yalla, and A. H. Mridul, “Optimizing e-commerce behavioral analytics: Strategy-driven ensemble blending,” International Journal of Advances in Computer Science & Engineering Research, vol. 1, no. 1, 2025.

[18] F. Weitao and Z. Chen, “Research on web crawler design and data collection method based on Python,” Smart City Appl., vol. 7, no. 12, pp. 123–125, 2024.

[19] M. J. Barwary, K. Jacksi, and A. Al-Zebari, “Constructing a multilingual e-learning ontology through web crawling and scraping,” Int. J. Commun. Networks Inf. Secur., vol. 15, no. 3, pp. 137–153, 2023.

[20] R. Anand, S. V. Lakshmi, D. Pandey, and B. K. Pandey, “An enhanced ResNet-50 deep learning model for arrhythmia detection using electrocardiogram biomedical indicators,” Evolving Syst., vol. 15, no. 1, pp. 83–97, 2024.