Proceedings of ICMRED 2026 Symposium: Financial Innovation, Risk Governance, and the Dynamics of Global Capital Flows
DOI: 10.54254/2754-1169/2026.BL32606

Research on the Construction of Sentiment Recognition and Analysis System for Financial Markets

Yiyang Bai

SWUFE-UD Institute of Data Science at SWUFE, Southwestern University of Finance and Economics, Chengdu, China
2646026227@qq.com

Abstract. Financial market sentiment has a crucial impact on market volatility and investment decisions, but existing research are limited by sufficient adaptability to multimodal data and incomplete sentiment analysis frameworks, which restrict their applicability in real-world financial scenarios. This paper proposes a targeted financial market sentiment analysis system that addresses the characteristics of multimodal financial data through dedicated data collection and preprocessing methods. A feature-level fusion framework based on a cross-modal attention mechanism is designed, and its effectiveness is validated through comparative experiments. Furthermore, a multi-level sentiment modeling architecture consisting of a base layer, intermediate layer, and application layer is constructed to support real-time sentiment classification, short-term prediction, and anomaly detection. On the basis, the system architecture, engineering implementation, and interface design of a sentiment analysis engine are completed. Experimental results demonstrate that the proposed fusion framework and hierarchical modeling system achieve superior performance in financial scenarios. Overall, this study overcomes the limitations of traditional single-modal sentiment analysis and provides a practical and scalable technical path for investment decision-making and risk management in fintech application.

Keywords: Multimodal data, Data fusion, Emotion modeling, Emotion Engine

1. Introduction

With the rapid development of big data and artificial intelligence technologies, multimodal data such as text, voice, and images have become new core information sources in financial market. Tese data carry rich market sentiment signals that reflect investors' expectations and market psychology, playing an irreplaceable role in understanding market dynamics and predicting price trends. However, due to the complexity and diversity of multimodal data, accurately capturing and analyzing financial market sentiment remains a key challenge.

In recent years, research on sentiment analysis in financial markets has advanced rapidly worldwide. International studies mainly focuses on applying sentiment analysis techniques to markets such as stocks and foreign exchange, developing a series of analytical models based on text, voice, and social network data. Domestic research, in contrast, emphasizes sentiment analysis frameworks that integrate multiple data sources and has achieved significant progress, particularly in

DOI: 10.54254/2754-1169/2026.BL32606

financial sentiment prediction and risk management. However, existing research still has several limitations: First, methods for extracting and preprocessing multimodal data features adapted to financial scenarios remain imperfect, and inconsistent data quality restricts analytical accuracy. Second, current multimodal fusion frameworks do not fully consider the specific characteristics of financial markets, making it difficult to capture complex emotional changes. Third, the construction of multi-level sentiment modeling systems for real-time classification, short-term prediction, and anomaly detection is immature, limiting the exploration of deeper sentiment value. Fourth, the construction and integration of sentiment engines that translate results into practical applications lag behind, lacking efficient output and application interfaces.

To address these gaps, this research aims to construct a financial market sentiment analysis system based on multimodal data fusion to enhance the accuracy and practicality of sentiment recognition and prediction. The research focuses on multimodal data collection, preprocessing, and fusion, as well as the construction of a multi-level sentiment modeling system, ultimately enabling the design and application of a sentiment engine. By incorporating advanced technologies such as cross-modal attention mechanisms and deep learning, and through experimental validation, the proposed system is expected to provide effective sentiment reference for financial decision-making.

2. Characteristics and preprocessing of multimodal data in financial scenarios

2.1. Types and characteristics of multimodal financial data

In financial market sentiment analysis, multimodal data refers to data from different channels and formats that can reflect the emotional states of market participants from multiple perspectives. Major types of multimodal financial data include text data, image data, audio data, and social media data. Text data typically originates from news reports, financial commentaries, and investor forums, and contains rich information about market dynamics and sentiment. For example, by analyzing keywords and sentiment trends in news reports, the emotional state of market participants can be inferred [1].

Image data may include stock market charts and graphical representations of economic indicators, which visually demonstrate market trends and fluctuations [2]. Audio data may include financial broadcasts, investor interviews, etc. By analyzing non-verbal information such as tone, pitch, and speech rhythm, market sentiment can be further understood [3]. Social media data, such as user posts on Weibo and Twitter, are an important source of real-time market sentiment. These data are usually highly real-time and have a wide social impact. To more intuitively show the types and characteristics of multimodal financial data, Figure 1 below shows the proportion of different types of data in financial market sentiment analysis [4].

DOI: 10.54254/2754-1169/2026.BL32606

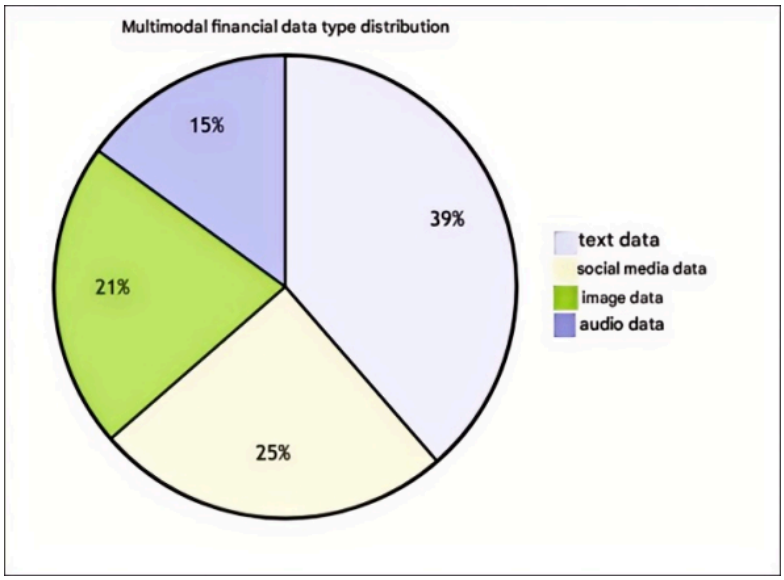

Figure 1. Distribution of multimodal financial data types

Furthermore, the processing of multimodal financial data must account for its diversity and complexity. For example, text data processing may require natural language processing techniques, while processing image data involves computer vision methods. Accordingly, preprocessing strategies should be designed based on the specific characteristics of each data type to ensure data quality and the effectiveness of the analysis.

2.2. Acquisition and screening of multimodal data

In building a financial market sentiment analysis system, the collection and screening of multimodal data is a fundamental and crucial step, directly impacting the effectiveness of subsequent data processing and model training. Multimodal data mainly includes text, images, and audio, each of which requires specific collection methods and screening criteria [5].

Text data is primarily obtained from public sources such as financial news, social media, and forums. Web crawling technology is used to automatically extract relevant text information, followed by initial cleaning using natural language processing (NLP) to remove irrelevant characters and stop words. Text data selection is typically based on keyword matching and sentiment analysis to ensure data relevance and quality. Image data mainly comes from financial news reports and social media platforms, and requires processing using computer vision techniques, including image denoising and feature extraction. During selection, the focus is on key information, such as economic indicator charts and market dynamics. Audio data, including financial interviews and podcasts, must first be converted into textual form through speech recognition technology before further analysis. Audio data selection emphasizes content relevance and clarity to ensure that the converted text data accurately reflects the original audio content. The data collection and selection process can be represented by Figure 2:

DOI: 10.54254/2754-1169/2026.BL32606

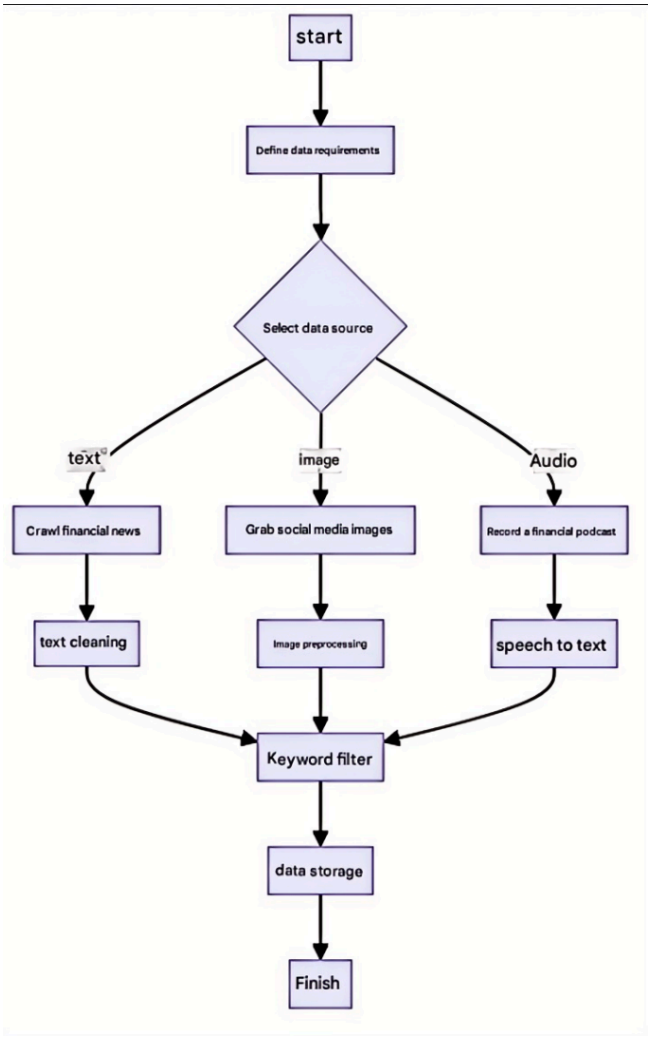

Figure 2. Data collection and filtering flowchart

Through the above process, high-quality multimodal data can be effectively collected and filtered from various sources, providing a solid data foundation for subsequent sentiment analysis.

2.3. Preprocessing methods for multimodal data

Multimodal data preprocessing is a crucial step in ensuring the effectiveness and efficiency of model training. This process typically includes data cleaning, standardization, normalization, and denoising. For financial data, preprocessing requires particular attention to data integrity and consistency, removing outliers and irrelevant features. Furthermore, time-series financial data often requires additional processing, such as sliding window methods and differencing, to reduce the influence of trends and seasonality. For unstructured data such as images, text, and audio, extracting key features and transforming them into formats suitable for machine learning algorithms is especially important. By comprehensively applying these strategies, the performance of subsequent fusion frameworks and sentiment analysis models can be significantly improved.

DOI: 10.54254/2754-1169/2026.BL32606

3. Design of a multimodal data fusion framework adapted to financial scenarios

3.1. Design principles of multimodal fusion framework

The design of the multimodal fusion framework follows five key principles. First, integration, ensuring that different types of financial data can be effectively integrated and processed within a single framework. Second, adaptability, requiring the framework to be flexibly respond to dynamic changes in financial markets and variations in data characteristics. Third, accuracy, emphasizing the use of advanced algorithms to improve the accuracy of sentiment recognition. Fourth, real-time performance, considering the high timeliness requirements of financial sentiment analysis, the system should support real-time data processing and analysis. Fifth, reliability, ensuring stable operation and consistent analytical performance when handling large-scale and complex financial data.

3.2. Overall system architecture

The overall architecture of the proposed system is designed to achieve efficient and accurate financial market sentiment analysis. It consists of five main layers: a data acquisition layer, a data processing layer, a multimodal fusion layer, a sentiment modeling layer, and a sentiment engine output layer.

The data acquisition layer collects raw data from various financial information sources, including news reports, social media, and financial forums. The data processing layer cleans, formats, and standardizes the collected data to ensure data quality. The multimodal fusion layer employs a cross-modal attention mechanism to integrate different financial data and extract comprehensive features [6]. Based on these fused features, the sentiment modeling layer constructs sentiment prediction models to identify and predict market sentiment trends [7]. Finally, the sentiment engine output layer transforms sentiment analysis results into actionable information to support financial decision-making. Through this layered architecture, the system ensures that final sentiment analysis results accurately reflect market dynamics and provide valuable decision support [8].

3.3. Feature-level fusion model based on cross-modal attention mechanism

In financial multimodal data analysis, different modalities convey different information. To effectively capture cross-modal correlations and improve sentiment prediction accuracy, this study proposes a feature-layer fusion model based on a cross-modal attention mechanism [9].

The core of the model is to automatically assign higher weights to modality features that contribute more significantly to sentiment prediction. Specifically, features from each modality are first encoded into high-dimensional vectors using modality-specific encoders. An attention module then computes correlation weights between feature representations from different modalities, and the weighted features are fused into a unified representation [10].

Formally, given feature vectors from multiple modalities, the cross-modal attention mechanism can be represented as:

$$w=\mathrm{softmax}(A(v_1,v_2,...,v_N))$$

Here, A is an attention function used to calculate the correlation between feature vectors of different modalities, w and is the weight vector adjusted by the attention mechanism. Next, we will demonstrate how to implement this attention mechanism using Python and the PyTorch framework:

DOI: 10.54254/2754-1169/2026.BL32606

```
Python
import torch
import torch.nn as nn
import torch.nn.functional as F
class CrossModalAttention(nn.Module):
def __init__(self, dim):
super(CrossModalAttention, self).__init__()
self.query = nn.Linear(dim, dim)
self.key = nn.Linear(dim, dim)
self.value = nn.Linear(dim, dim)
def forward(self, v_list):
# v_list is a list of feature vectors from different modalities
       q = self.query(v_list[0])
       k = self.key(v_list [1])
       v = self.value(v_list [2])
       # Compute attention scores
       attn_scores = torch.matmul(q, k.transpose(-2, -1)) / (k.size(-1)**0.5)
attn_weights = F.softmax(attn_scores, dim=-1)
       # Apply attention to the value vector
       weighted_v = torch.matmul(attn_weights, v)
       return weighted_v
# Example usage
dim = 128
model = CrossModalAttention(dim)
features = [torch.randn(1, dim), torch.randn(1, dim), torch.randn(1, dim)]
output = model(features)
```

This module dynamically adjusts feature weights across modalities based on their relevance, enabling more effective multimodal feature fusion and improving the accuracy of sentiment analysis.

3.4. Comparative verification of the fusion framework

To evaluate the effectiveness of the proposed multimodal data fusion framework, a series of comparative experiments were conducted. The proposed method was compared with several mainstream fusion methods under the same experimental settings [11].

DOI: 10.54254/2754-1169/2026.BL32606

Table 1. Validation of the multimodal data fusion framework

Method	Accuracy (%)	Recall rate (%)	F1 score (%)
Method A	82	78	80
Method B	85	80	82.5
Method C	79	76	77.5
Proposed method	90	88	89

As shown in Table 1, the proposed framework achieves superior performance in terms of accuracy, recall, and F1 score. Figure 3 further illustrates the overall performance distribution of different fusion methods.

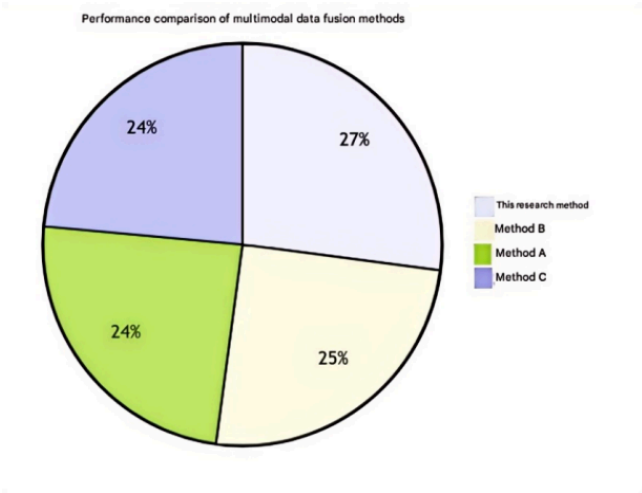

Figure 3. Performance comparison of multimodal data fusion methods

These results demonstrate that the proposed multimodal data fusion framework is more effective in handling complex financial data and capturing market sentiment, thereby providing a solid foundation for subsequent sentiment modeling and application research [12].

4. Construction of a multi-level financial market sentiment modeling system

Market sentiment modeling plays a crucial role in financial analysis by capturing investors' emotional responses and their impact on market trends and volatility. To addresses the heterogenity and complexity of sentiment data, this study proposes a multi-level sentiment modeling system, consisting of the foundational layer, the intermediate layer, and the application layer.

The foundational layer focuses on the accurate identification and classification of real-time emotions. It processes textual, visual, and audio information collected from financial views, social media platforms, and investor commentaries. Deep learning techniques, particularly NLP and computer vision, are employed to extract emotional features and sentiment indicators [13].

Models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory networks (LSTMs) are used to capture semantic and temporal features in financial discourse. By learning from large-scale domain-specific data, the model adapts to the distinctive emotional expressions of financial markets. To accommodate the dyynamic nature of market sentiment, the system is designed with real-time updating capabilities, enabling timely sentiment recognition and classification [14].

DOI: 10.54254/2754-1169/2026.BL32606

Building upon the outputs of the foundational layer, the intermediate layer aims to predict short-term sentiment trends in financial markets. This layer integrates historical sentiment patterns with real-time signals to model temporal dependencies in sentiment evolution.

Time-series models such as RNNs and LSTMs are applied to capture short-term fluctuations, while attention mechanisms emphasize key information. Model parameters are optimized using historical market data to ensure prediction accuracy and generalization ability. This layer enables early identification of sentiment shifts that may precede market movements.

The application layer is responsible for detecting abnormal or extreme sentiment fluctuations that may indicate heightened market risk. Using sentiment sequences generated by previous layers, the model analyzes deviations from historical sentiment patterns.

LSTM-based time-series analysis combined with sliding window techniques is employed to monitor sentiment dynamics within specific time intervals. Additionally, anomaly detection algorithms—such as Isolation Forest and statistical outlier detection—are integrated to improve robustness. This layer supports timely identification of sentiment shocks and enhances market risk monitoring.

5. Construction and integration of financial market sentiment engine

The financial market sentiment engine adopts a modular architecture, consisting of a data acquisition layer, a multimodal processing layer, a sentiment model layer, and an output application layer. The data acquisition layer collects information from multiple sources, including social media, news reports, and stock market trading data. The preprocessing layer ensures data consistency and quality across different modalities. The sentiment model layer integrates financial domain knowledge with machine learning models to generate sentiment indicators. Finally, the application layer transforms analytical results into sentiment indices, trend visualizations, and early warning signals, supporting practical decision-making.

Key system modules include data processing, feature extraction, model training, and sentiment analysis. In data processing, distributed computing techniques are employed to support efficient data collection and cleaning. Deep learning models are used for multimodal feature extraction, while model training is conducted on high-performance computing platforms to ensure stability and scalability. Through coordinated module interaction and standardized engineering workflows, the sentiment engine achieves reliable real-time analysis of financial market sentiment.

The sentiment engine provides outputs such as sentiment indices, fluctuation curves, and short-term trend predictions. To support system integration, the engine offers standardized output formats, such as JSON and RESTful API interactions, enabling seamless connection with financial analysis platforms. Access control mechanisms are incorporated to ensure data security and analytical reliability. The interface design emphasizes scalability and extensibility, facilitating future model upgrades and feature iterations.

6. Experimental verification and effect analysis

6.1. Experimental data and environment

The experiments were conducted using publicly available datasets from various financial markets, including stocks, foreign exchange, and cryptocurrencies. The data sources consist of historical tading records, financial news articles, and social media texts, ensuring both diversity and representativeness of market sentiment information. All datasets underwent standardized

DOI: 10.54254/2754-1169/2026.BL32606

preprocessing proceduresm including data cleaning, standardization, and feature extraction techniques to guarantee data quality. The experimental environment was configured with high-performance GPU servers to support complex multimodal data processing and deep learning model training tasks. System development and experimental verification primarily relied on the Python programming language and related deep learning frameworks such as TensorFlow and PyTorch, ensuring the efficiency and accuracy of the algorithm implementation.

6.2. Performance verification of the multimodal fusion framework

The study tested the multimodal fusion framework using real financial data. By comparing and analyzing the model's performance under different conditions, the practicality and efficiency of the fusion framework were effectively verified. Evaluation metrics included accuracy, recall, and F1 score. The experimental results show that the designed cross-modal attention mechanism model exhibits high accuracy in financial sentiment prediction tasks, effectively capturing multimodal information in the financial market and accurately predicting market sentiment trends. Furthermore, comparison with other fusion strategies further highlights the advantages of this framework in improving the accuracy and efficiency of sentiment prediction. In conclusion, this multimodal fusion framework demonstrates significant effectiveness in financial market sentiment analysis.

6.3. Validation of the emotion modeling system

In evaluating the proposed multi-level financial market sentiment modeling system, a combination of quantitative and qualitative analysis methods was adopted. Experimental results show that the accuracy of the basic layer real-time sentiment classification model reaches 92%, an improvement of nearly 10% points compared to traditional models.

The short-to-medium-term sentiment trend prediction model maintains a prediction error rate within 5%, indicating its high accuracy in capturing sentiment fluctuation trends. The abnormal sentiment detection model effectively identifies extreme sentiment deviations, providing decision support for investors. The overall effectiveness verification of this system demonstrates that it can provide a scientific and accurate tool for financial market sentiment analysis, helping investors make more rational investment decisions in a complex and volatile financial environment.

6.4. Practical application validation of the emotion engine

The practical applicability of the sentiment engine was further evaluated using real financial market data. Historical data were used for model training, while recent market data were employed for testing to ensure objectivity and practicality of the evaluation results.

Comparative experiments with traditional sentiment analysis methods demonstrate the system's advantages and its adaptability and effectiveness in the financial market. Through comprehensive analysis, this paper will elaborate on the contribution of the sentiment engine to improving the accuracy of financial market sentiment analysis [15].

7. Conclusion

This study develops a multimodal data fusion system for financial market sentiment analysis by integrating multi-source data such as text, images, and videos. The proposed cross-modal attention mechanism demonstrates strong performance in feature-layer fusion, enabling effective capture of sentiment correlations across different modalities. Furthermore, the multi-level sentiment modeling

DOI: 10.54254/2754-1169/2026.BL32606

system, progressing from real-time classification to abnormal sentiment detection, provides market participants with a powerful sentiment prediction tool. Experiment results confirm the effectiveness and practicality of the proposed system in real-world financial scenarios, demonstrating its potential value for financial decision support.

Despite these contributions, several limitations remain: constraints in data source may affect the generalization ability of the modal across different financial markets; the applicability of the multimodal data fusion framework to specific financial markets needs further verification; and real-time processing performance may be challenged when dealing with large-scale data. Furthermore, sentiment analysis itself still faces inherent difficulties, such as accurately identifying complex linguistic phenomena like irony and metaphor.

Future research will focus on improving the system's responsiveness to sudden events and enhancing real-time performance under large-scale data conditions. Further efforts may include optimizing sentiment prediction algorithms, strengthening the robustness of abnormal sentiment detection models, and incorporating additional unstructured data sources to enrich sentiment representation. These improvements are expected to enhance the accuracy, stability, and practical value of multimodal financial market sentiment analysis systems.

References

[1] Li, H.L, Ren, C.S., Liu, X.R., et al. (2023) A review of textual sentiment research in financial markets. Data analysis and knowledge discovery (12): 22-39.
[2] Wang, R.S.Y. (2021) Research on Stock Market Sentiment and Trend Analysis Model Based on Deep Learning. Huazhong University of Science and Technology. DOI: 10.27157/d.cnki.ghzku.2021.005348.
[3] Chen, H. (2020) Investor Sentiment Analysis and Application Based on Deep Learning. China Jiliang University. DOI: 10.27819/d.cnki.gzgjl.2020.000419.
[4] Shen, C., Jiang, Z.W., Cheng, D.L., et al. (2019) Construction and Application Analysis of Sentiment Index in Network Big Data—Taking the Securities Market as an Example. Wireless Interconnection Technology, (15): 11-12.
[5] Jiao, L. (2020) Design and Implementation of a Multimodal Sentiment Analysis System. Beijing University of Posts and Telecommunications. DOI: 10.26969/d.cnki.gbydu.2020.000943.
[6] Ding, J., Yang, L., Lin, H.F., et al. (2022) Research on Sentiment Analysis Based on Multimodal Heterogeneous Dynamic Fusion. Journal of Chinese Information Processing. (05): 112-124.
[7] Chen, S.H. (2022) Multimodal Fusion Sentiment Analysis Based on Deep Learning. Hangzhou Dianzi University.
[8] Xu, T.J. (2019) Investor sentiment and stock price volatility from a behavioral finance perspective. Guangxi Quality Supervision Guide (05): 198.
[9] Geng, Y.W. (2022) Design and Implementation of Multimodal Sentiment Analysis System. Beijing University of Posts and Telecommunications. DOI: 10.26969/d.cnki.gbydu.2022.002879.
[10] Zou, J.Y. (2020) Multi-level modal representation fusion for sentiment analysis. Hebei University of Science and Technology, 2020. DOI: 10.27107/d.cnki.ghbku.2020.000232.
[11] Qiu, S.Y. (2022) Exchange rate forecasting based on a multi-factor, multi-scale intelligent optimization deep framework incorporating investor sentiment. Nanjing University of Information Science and Technology. DOI: 10.27248/d.cnki.gnjqc.2022.000348.
[12] Pan, J.Y. (2021) Construction and Validation of Financial Market Sentiment Index Based on Text Mining. Hangzhou Dianzi University. DOI: 10.27075/d.cnki.ghzdc.2021.000059.
[13] Li, F.W. (2022) Research on Deep Reinforcement Learning Stock Trading Model Combined with Sentiment Analysis. Guangdong University of Finance and Economics. DOI: 10.27734/d.cnki.ggdsx.2022.001046.
[14] Li, D. (2023) Design and Implementation of Emotion Recognition System Based on Multimodal Fusion. Shenyang University of Technology. DOI: 10.27322/d.cnki.gsgyu.2023.001380.
[15] Gao, J.H. (2022) Research on Quantitative Trading Strategies Based on Investor Sentiment. Guangdong University of Finance and Economics. DOI: 10.27734/d.cnki.ggdsx.2022.000196.