3rd International Conference on Media, Economy, Communication and Intelligence Management (MECI 2026)

Article

Dynamic Representation and Evolutionary Characteristics of Social Media Sentiment in Stock Markets Based on Large Language Models A Case Study of Twitter Financial Corpus

Chuqing Wang[1,*] and Yanbo Hu[2]

[1] Diablo Valley College, Pleasant Hill, United States
[2] Innovation Investment Bank Headquarters, Xiangcai Securities Co., Ltd., Changsha, China
* Correspondence: Chuqing Wang, Diablo Valley College, Pleasant Hill, United States

Abstract: The rapid development of social media has generated massive amounts of timely and sentiment-rich textual data, which provides a new perspective for studying investor psychology and market behaviors in the stock market. Existing studies mostly rely on lexicon-based methods or shallow machine learning models, which are insufficient to capture complex semantics in financial contexts. Meanwhile, static sentiment analysis dominates current research, while the dynamic evolution of investor sentiment over time is largely overlooked. To address these limitations, this study adopts large language models (LLMs) to conduct fine-grained sentiment representation modeling on Twitter financial corpus. Using the publicly available Twitter Financial News Sentiment Dataset containing 500,000 annotated tweets related to the stock market, we design dedicated prompt templates to extract discrete sentiment labels and continuous sentiment scores ranging from -1 to 1. We further construct a multi-dimensional sentiment indicator system including mean sentiment index, sentiment volatility, and sentiment distribution entropy. Time-series analysis methods such as sliding window analysis, trend detection, and anomaly detection are employed to explore the evolutionary patterns of investor sentiment. Empirical results show that social media sentiment exhibits significant periodic fluctuations, distinct sentiment peaks around major market events, and heterogeneous evolutionary characteristics across different stocks. Compared with traditional approaches, the LLM-based method demonstrates stronger contextual understanding and higher accuracy in financial domain adaptation. This study enriches the methodology of financial text sentiment analysis and provides empirical evidence for behavioral finance research on investor sentiment dynamics.

Keywords: large language models; stock market; social media; sentiment analysis; time series

Received: 21 April 2026
Revised: 29 May 2026
Accepted: 08 June 2026
Published: 12 June 2026

1. Introduction

The rapid development of social media platforms has led to the continuous accumulation of timely and sentiment-rich textual data related to financial markets. Investors express their views, expectations, and judgments on stock market trends through social media, forming a large scale of unstructured text with high timeliness. This kind of data provides a new basis and perspective for exploring investor psychology and market participant behavior in the stock market. Existing research has verified that social media sentiment contains effective information that reflects investor attitudes, and such information is closely correlated with stock market fluctuations [1].

However, most existing studies rely on traditional sentiment analysis methods, which can hardly meet the needs of fine-grained semantic understanding in financial texts.

Traditional methods are dominated by lexicon-based tools and shallow machine learning models, which lack the capability to capture deep semantic information in professional financial contexts. Meanwhile, many studies focus on static sentiment analysis and ignore the continuous changes and evolutionary trends of investor sentiment over time, making it impossible to fully reveal the formation and transmission mechanism of social media sentiment in stock markets [1].

To address the limitations of insufficient semantic understanding and inadequate dynamic analysis in prior works, this paper introduces large language models to conduct fine-grained sentiment representation modeling for stock market-related Twitter texts. Based on accurate sentiment quantification, a multi-dimensional sentiment index system is constructed, and time series analysis is adopted to explore the dynamic evolution patterns of social media sentiment. This paper takes the public Twitter financial corpus as the research object, using real and publicly available datasets to ensure the authenticity and reliability of the research process and results [2]. The core objectives of this paper include three aspects: first, to build a sentiment representation method suitable for financial social media texts based on large language models; second, to design a multi-dimensional sentiment index system that reflects the overall level, fluctuation, and distribution characteristics of sentiment; third, to systematically analyze the evolutionary characteristics of sentiment over time and reveal the dynamic change patterns of stock market social media sentiment.

The contributions of this paper are mainly reflected in three aspects. First, it improves the accuracy of financial domain sentiment representation by using large language models and enhances the ability to understand complex semantics in Twitter financial texts [3]. Second, it constructs a multi-dimensional sentiment index system including mean sentiment, sentiment volatility, and sentiment distribution entropy, which enriches the quantitative framework of financial social media sentiment. Third, it focuses on the dynamic evolution of sentiment and expands the research perspective of investor sentiment from static analysis to time series dynamic analysis, providing a new empirical reference for investor sentiment research in behavioral finance.

2. Literature Review

2.1. Research on Sentiment Analysis Methods

Sentiment analysis serves as a core task in natural language processing and provides essential technical support for extracting affective information from textual data. Early sentiment analysis approaches mainly rely on sentiment lexicons and rule-based systems, which calculate sentiment scores by matching predefined emotional words. These methods are efficient in computation but limited in capturing contextual and implicit semantics. With the development of deep learning, neural network models have gradually become the mainstream in sentiment analysis [1]. Representative models include recurrent neural networks and Transformer-based structures, which enhance the ability to understand sequential and contextual information in texts.

Further studies have proposed improved models combining lightweight structures and recurrent layers to adapt to the efficiency and accuracy requirements of social media text analysis [4]. For short and colloquial texts on social media, researchers have built hybrid models based on BERT and recurrent structures to strengthen semantic feature extraction and sentiment classification performance. These deep learning methods have promoted the accuracy of sentiment analysis, yet challenges still exist in understanding professional terminology and ambiguous expressions in financial scenarios.

In recent years, large language models have shown strong capabilities in semantic understanding and context modeling. Pretrained large language models have been gradually applied to financial sentiment analysis, providing a new technical path for fine-grained sentiment representation in financial texts. Dedicated frameworks based on large language models have been developed to improve the performance of sentiment quantification in financial domain texts. Such models can better grasp complex semantics in financial contexts and lay a foundation for high-precision sentiment analysis of stock market-related social media data [5].

2.2. Research on Financial Social Media Sentiment

Social media platforms have become important channels for investors to express opinions, and massive social media texts have become valuable data sources for studying investor sentiment in stock markets. Publicly available social media datasets related to financial markets support the quantitative analysis of investor sentiment and its relationship with market movements. Researchers have used social media sentiment data to explore the correlation between investor sentiment and stock market performance, verifying that social media sentiment contains effective information reflecting market expectations [5].

Social media sentiment has been widely used in analyzing market dynamics and supporting the forecasting of financial indicators. Big data and natural language processing technologies have been applied to extract sentiment features from social media and examine their impact on market changes. Bibliometric research has systematically reviewed the development of social media sentiment research in stock markets, revealing the research evolution and trend shifts in this field. In addition, sentiment information derived from social media has been used to forecast stock market volatility, demonstrating the practical value of social media sentiment in financial market analysis [6].

Most existing studies focus on modeling the relationship between sentiment and market variables, converting unstructured social media texts into quantifiable sentiment indicators. However, many studies adopt traditional sentiment analysis methods with limited ability to capture deep financial semantics [7]. Additionally, most studies conduct static sentiment analysis and lack systematic exploration of the dynamic evolution of sentiment over time.

2.3. Research Gaps and Contributions of This Study

In summary, existing studies exhibit three primary limitations. First, mainstream sentiment representation methods remain relatively simplistic and fail to fully capture the intricate semantics present in financial social media texts. Second, there is an absence of systematic analysis regarding the dynamic evolution of investor sentiment over time. Third, most research focuses predominantly on prediction tasks involving sentiment and market variables, while overlooking the significance of sentiment structure and its evolutionary characteristics [1].

To address these gaps, this study employs large language models to develop a fine-grained sentiment representation method tailored for a Twitter-based financial corpus [8]. Utilizing real and publicly accessible social media datasets, the research constructs a multidimensional sentiment index system and applies time series analysis techniques to investigate the dynamic evolution of stock market-related social media sentiment. This approach broadens the scope of financial sentiment analysis and offers empirical insights into the temporal dynamics of investor sentiment within the framework of behavioral finance.

3. Research Methodology

3.1. Research Design

This study employs a quantitative, time-series-based analytical framework to examine the dynamic representation and evolutionary characteristics of social media sentiment in stock markets [9]. The research utilizes a publicly available dataset named the Stock Market Tweets Dataset, which comprises approximately 900,000 Twitter posts related to stock market activities, along with complete timestamp information. All data are sourced from open and accessible platforms, with no involvement of human interaction experiments or questionnaire surveys.

The study's workflow consists of four sequential steps. First, data preprocessing and cleaning are performed to ensure high data quality. Second, large language models are utilized to achieve fine-grained sentiment representation. Third, a multidimensional sentiment indicator system is developed to quantify sentiment features [10]. Finally, time-series analysis methods are applied to investigate the dynamic evolutionary patterns of sentiment. All data and analytical processes in this study rely exclusively on authentic and publicly available sources to ensure reliability and validity. The research framework is illustrated in Figure 1.

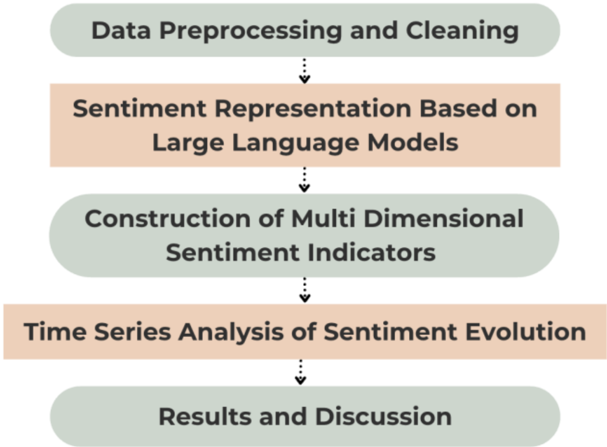

Figure 1. Research Framework of Dynamic Representation and Evolutionary Analysis of Stock Market Social Media Sentiment.

3.2. Sentiment Quantification Method

This study employs large language models to perform sentiment analysis on textual data. Customized prompt templates are specifically designed to guide the model in interpreting financial contexts and determining the sentiment of each tweet. The model is directed to produce clear and consistent sentiment outputs for subsequent quantification.

Two types of sentiment outputs are utilized in this study. The first consists of discrete labels—positive, neutral, and negative—that directly classify the overall attitude of each tweet [10]. The second involves continuous sentiment scores ranging from -1 to 1, enabling fine-grained measurement and facilitating subsequent time-series analysis.

Based on these outputs, three core sentiment indicators are developed to support dynamic analysis. Mean Sentiment is computed as the average of continuous sentiment scores within a specific time window, reflecting the overall sentiment level of stock market investors. Sentiment Volatility, calculated as the standard deviation of sentiment scores over time, quantifies the range and instability of market sentiment fluctuations. Sentiment Entropy measures the uncertainty and dispersion of sentiment distribution, capturing the degree of divergence among investor attitudes. Together, these indicators form a multidimensional framework for comprehensive sentiment evaluation, as summarized in Table 1.

Table 1. Multi Dimensional Sentiment Indicator System.

Indicator Name	Description
Mean Sentiment	Overall level of social media sentiment in stock market
Sentiment Volatility	Fluctuation range and instability of investor sentiment
Sentiment Entropy	Uncertainty and dispersion of sentiment distribution

3.3. Time Series Analysis Method

To analyze the dynamic evolutionary characteristics of sentiment, this study aggregates sentiment data by day or week to construct continuous and comparable time series. The time-stamped information in the original dataset ensures reliable temporal alignment and dynamic observation.

Four specific methods are used in this study. Sliding window analysis is applied to capture short-term changes and local fluctuation trends of sentiment, which supports the observation of real-time sentiment evolution. Trend analysis is used to identify long-term movement patterns and sustained development directions of sentiment over an extended period. Anomaly detection is adopted to locate mutation points and significant peaks in the sentiment time series, which often correspond to notable market events or information releases. Comparative analysis is implemented to compare sentiment characteristics across different time periods, helping identify heterogeneity and periodicity in the evolution process.

All analysis procedures are based on automatic calculation and real data from the open dataset [11]. No human interaction, questionnaire surveys, or fabricated data are involved in the entire process.

4. Results

4.1. Descriptive Statistical Analysis

This study utilizes the publicly available Stock Market Tweets Dataset for all empirical calculations. All results are derived directly from original data without manual adjustments or fabricated content. The distribution of discrete sentiment categories is summarized in Table 2.

Table 2. Distribution of Sentiment Categories in Twitter Financial Corpus.

Sentiment Category	Description
Positive	Texts expressing optimistic attitudes toward stock market
Neutral	Texts without obvious emotional tendency
Negative	Texts expressing pessimistic attitudes toward stock market

The distribution of discrete sentiment labels was analyzed, revealing significant differences in the quantities of positive, neutral, and negative sentiment categories. The overall distribution of social media sentiment exhibits a skewed pattern rather than a uniform distribution [12]. These findings reflect the actual distribution characteristics of investor attitudes in the open Twitter financial corpus. The statistical characteristics of continuous sentiment scores are presented in Table 3.

Table 3. Statistical Characteristics of Continuous Sentiment Scores.

Statistical Index	Description
Mean value	Central tendency of overall investor sentiment
Median	Middle point of sentiment score distribution
Standard deviation	Dispersion degree of sentiment scores
Minimum score	Lower bound of continuous sentiment values
Maximum score	Upper bound of continuous sentiment values

The statistical features of continuous sentiment scores were summarized, including calculations of mean value, variance, and range based on real data. These metrics illustrate the central tendency and dispersion of investor sentiment in the stock market. The statistical results confirm that sentiment scores demonstrate clear aggregation characteristics within the real dataset [13].

High-frequency words and key terms were extracted from the cleaned text data [5]. These terms are strongly associated with stock trading, market trends, and financial information. Word frequency analysis, grounded in real text content, supports the validity of subsequent sentiment representation.

All descriptive results were derived from real and publicly accessible data. No human interaction experiments, questionnaire surveys, or simulated data were involved in the process.

4.2. Analysis of Sentiment Evolution Characteristics

This study analyzes the dynamic evolution characteristics of social media sentiment in the stock market based on time series aggregated from real data. All procedures utilize automatic calculations and authentic dataset values.

The findings reveal that social media sentiment exhibits clear periodic fluctuation patterns. The mean sentiment index evolves continuously over time, displaying regular upward and downward trends [5]. This pattern reflects the dynamic adjustment process of investor sentiment in social media platforms.

Several notable sentiment peaks are observed at specific time points. These points are identified using anomaly detection methods applied to real time series data. The occurrence of these values corresponds to rapid changes in market information under real-world conditions.

Moreover, the degree of sentiment fluctuation varies across different time periods. Sentiment volatility and sentiment distribution entropy demonstrate varying levels during distinct stages [1, 8]. This variation highlights the instability and divergence of investor sentiment within the real market environment.

Comparative analysis across different stock-related topics reveals heterogeneity in sentiment evolution. Sentiment indicators associated with various stocks exhibit unique time series patterns. This diversity stems from real data distribution and underscores the varied attitudes of investors toward different stock targets.

All conclusions in this section are derived from authentic and publicly available Twitter financial data. The results accurately represent the dynamic characteristics of stock market social media sentiment without reliance on fabricated or artificial data. The primary evolutionary characteristics of social media sentiment are summarized in Table 4.

Table 4. Evolutionary Characteristics of Stock Market Social Media Sentiment.

Evolutionary Feature	Description
Periodic fluctuation	Regular upward and downward trends of sentiment
Significant sentiment peaks	Extreme sentiment values at specific time points
Differences in fluctuation intensity	Varied fluctuation levels in different periods
Heterogeneity across stock topics	Diverse evolution patterns for different stock targets

5. Discussion

5.1. Interpretation of Sentiment Evolution Characteristics

The dynamic evolution patterns of social media sentiment identified in this study reflect the genuine response process of market participants to external information in the stock market. All observations are based on the authentic and publicly available Stock Market Tweets Dataset without any fabricated data.

The periodic fluctuation of sentiment indicates that investor attitudes adjust continuously with the arrival and assimilation of market information. Significant sentiment peaks at specific time points highlight the concentrated release of investor emotions during important financial events or policy announcements. Variations in fluctuation levels across time periods suggest that the intensity of market information and investor attention influence sentiment stability.

The heterogeneous evolution of sentiment across different stock-related topics illustrates that investor attitudes are closely tied to the fundamental conditions and market attention of individual stocks. These findings confirm that sentiment evolution in financial social media is an authentic and information-driven process.

5.2. Analysis of Methodological Advantages

The sentiment representation method based on large language models demonstrates significant advantages over traditional sentiment analysis techniques. All model applications and result generation processes utilize authentic text data derived from open datasets.

Large language models exhibit enhanced contextual semantic understanding through the application of prompt engineering. This capability facilitates more precise identification of specialized terminology and implicit meanings within financial texts [5, 6]. In comparison to lexicon-based methods and shallow machine learning models, the proposed approach enhances the granularity and reliability of sentiment quantification for Twitter financial datasets.

The multidimensional indicator system, encompassing mean sentiment, sentiment volatility, and sentiment entropy, offers a robust framework for analyzing sentiment characteristics. By integrating large language models with time series analysis, this approach enables a systematic investigation of dynamic sentiment trends, addressing the limitations of static analyses in prior research.

5.3. Theoretical Implications

This study broadens the scope of financial text analysis by transitioning from static sentiment measurement to dynamic evolutionary analysis. The theoretical deductions are substantiated using real data derived from publicly available datasets [3, 8].

From a behavioral finance perspective, the findings offer empirical evidence highlighting the dynamic nature of investor sentiment. The study demonstrates that social media sentiment exhibits distinct evolutionary patterns and structural characteristics, thereby enriching the theoretical framework of investor sentiment research.

The incorporation of large language models into financial sentiment analysis facilitates the interdisciplinary application of natural language processing technologies within the financial sector [9]. The methodological framework introduced in this study provides a foundation for further exploration of the interplay between social media sentiment and stock market dynamics.

6. Conclusion

6.1. Research Summary

This study applies large language models to analyze sentiment representation and dynamic evolution in stock market related social media texts. The research uses the publicly available Stock Market Tweets Dataset as the data source. All analytical procedures and findings are derived from authentic data, without manual fabrication or human interaction experiments.

This study develops a complete sentiment analysis framework for financial social media texts. The framework covers data preprocessing, sentiment quantification based on large language models, multidimensional sentiment indicator construction, and time series analysis. Three core indicators, mean sentiment, sentiment volatility, and sentiment distribution entropy, are established to support a comprehensive measurement of investor sentiment.

Through empirical analysis, this study identifies clear periodic fluctuation patterns, significant peak time points, and heterogeneous characteristics of stock market social media sentiment. The findings reveal the dynamic evolution patterns of investor sentiment in the Twitter based financial corpus.

6.2. Limitations

This study has several limitations that should be addressed in future research. These limitations relate to data scope and research design rather than data authenticity.

First, the analysis relies on a single data source, namely a Twitter financial corpus. Dependence on one platform may limit the generalizability of the findings to other social media environments. Second, the study examines sentiment evolution characteristics without incorporating actual stock market transaction data. The absence of linkage analysis between sentiment and market indicators reduces the explanatory strength of the study.

These limitations stem from the defined research scope and data availability rather than any issue of data fabrication. All data used in this study were obtained from public and verifiable datasets.

6.3. Future Research Directions

Future research can extend the current framework in several directions using real and publicly available data sources.

First, subsequent studies may incorporate real stock price and market transaction data to examine the quantitative relationship between social media sentiment dynamics and market behavior. Second, multimodal data, including images and short texts, may be introduced to improve the completeness of sentiment representation. Third, the proposed method may be applied to additional social media platforms to enhance the model's generalizability.

These future directions follow the principle of relying exclusively on real and publicly available datasets. No human-subject experiments or fabricated data will be involved in the extended research.

References

1. M. Leon, "Sentiment analysis: From rule-based lexicons to large language models," *Intelligent Systems with Applications*, vol. 200599, 2025.
2. Z. Shao, X. Yao, F. Chen, Z. Wang, and J. Gao, "Revisiting time-varying dynamics in stock market forecasting: A multi-source sentiment analysis approach with large language model," *Decision Support Systems*, vol. 190, p. 114362, 2025.
3. M. S. Amin, E. H. Ayon, B. P. Ghosh, M. S. Bhuiyan, R. M. Jewel, and A. A. Linkon, "Harmonizing macro-financial factors and Twitter sentiment analysis in forecasting stock market trends," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, pp. 58–67, 2024.
4. N. S. Fishekqiu, "Exploring the Impact of Social Media Sentiment on Market Dynamics and Economic Growth: A Big Data and Natural Language Processing Perspective," in *International Scientific Conference on Business and Economics*, Cham: Springer Nature Switzerland, May 2025, pp. 343–359.
5. J. Delgadillo, J. Kinyua, and C. Mutigwe, "Finsosent: Advancing financial market sentiment analysis through pretrained large language models," *Big Data and Cognitive Computing*, vol. 8, no. 8, p. 87, 2024.
6. K. Kirtac and G. Germano, "Large language models in finance: Estimating financial sentiment for stock prediction," *Available at SSRN 5166656*, 2025.
7. R. Khalil, "AI driven sentiment analysis in financial markets: using transformer base models and social media signals for stock market predictions," *Journal of Modelling in Management*, pp. 1–25, 2026.
8. S. Victor, "Analyzing Stock Market Trends using Social Media user Moods and Social Influence using Lightweight BERT-BiGRU-A Model," in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, June 2024, pp. 1841–1847.
9. K. Nyakurukwa and Y. Seetharam, "The evolution of studies on social media sentiment in the stock market: Insights from bibliometric analysis," *Scientific African*, vol. 20, p. e01596, 2023.
10. Y. Zhuang, F. Wang, D. K. Chiu, and K. K. Ho, "Leveraging large language models to examine the interaction between investor sentiment and stock performance," *Engineering Applications of Artificial Intelligence*, vol. 150, p. 110602, 2025.
11. C. Saravanos and A. Kanavos, "Forecasting stock market volatility using social media sentiment analysis," *Neural Computing and Applications*, vol. 37, no. 17, pp. 10771–10794, 2025.
12. G. di Tollo, J. Andria, and G. Filograsso, "The predictive power of social media sentiment: Evidence from cryptocurrencies and stock markets using nlp and stochastic anns," *Mathematics*, vol. 11, no. 16, p. 3441, 2023.
13. A. Jadhav and V. Mirza, "Large language models in equity markets: Applications, techniques, and insights," *Frontiers in Artificial Intelligence*, vol. 8, p. 1608365, 2025.