Beyond Single-Text Analysis: A Holistic Approach to Chinese Financial Sentiment

Mingfei Zhang[1], Chengze Du[2,✉]
[1] School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, China
[2] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China
Mingfeizhang05@gmail.com, ducz0338@bupt.edu.cn

***Abstract*—Financial sentiment analysis is crucial to comprehend the functioning of markets, particularly in China, where social media significantly influences the behavior of investors. Despite the fact that sentiment analysis has progressed significantly, analyzing Chinese financial social media content remains challenging due to the lack of labeled datasets and the sophistication of Chinese social media discussions, making sentiment analysis challenging and inaccurate. Conventional approaches struggle to comprehend the subtle context, such as sarcasm and new Internet slang, and fail to fuse various sentiments from social media interactions effectively. This paper proposes a framework [1] that addresses these issues by fusing context-aware designs and integrating various sentiment signals. Our approach grasps the layered context of conversations and integrates various measurements and their interrelations into one sentiment indicator. Experiments on Chinese financial news and social media data show that our system performs better than the traditional method on both sentiment classification tasks and downstream financial applications. This provides a stronger solution for Chinese financial sentiment analysis.**

***Index Terms*—Financial Sentiment Analysis, Chinese Social Media, Stock Forecast**

I. Introduction

The rapid development of social media websites has made a difference the world of financial markets, particularly in China where websites such as Eastmoney are crucial in sensing market emotions and making investment decisions. Automatic collection and research of public opinions from social media comments through financial sentiment analysis have emerged as an integral tool for market participants. Investors can discover more about market patterns, forecast market direction, and make smarter investment choices by studying sentiment on social media. Businesses use sentiment analysis as well to track public opinion, assess risks, and change their path accordingly. Recent methods in natural language processing and deep learning [11], [23] have permitted the deployment of more advanced sentiment analysis techniques. Still, most of these developments target English material. There are certain difficulties with Chinese financial sentiment analysis that need for certain remedies.

[1]Source code is available at https://gitee.com/Monickar/Fin-Sentiment-LLM.

Fig. 1. This figure illustrates the challenges of sentiment analysis in the Chinese stock market context. The left panel shows a news article about tariff policies and A-shares with accompanying comments, presented in both Chinese and English. The right panel demonstrates how sentiment classification can differ between single-text analysis and context-aware analysis incorporating Chinese memes and cultural nuances. The sentiment matrix shows five commented examples (#1-#5) where direct text analysis often yields different results from contextual analysis. For instance, comment #2's apparent positive sentiment ("genius") becomes negative when considering Chinese conversational irony. Similarly, the seemingly neutral comment #5 about "controlling hands" carries negative market implications in Chinese trading context. The diverging arrows on the right emphasize these sentiment shifts between surface-level and context-aware analysis.

The scarcity of labeled Chinese financial datasets hampers the development of robust models, while the inherent complexity of Chinese social media discourse—including sarcasm, Internet slang, and rapidly evolving memes—makes sentiment detection particularly difficult and error-prone. For instance, when retail investors comment "高手在民间" (genius among the masses) on a news post about significant market losses, traditional sentiment models might incorrectly interpret this as positive sentiment due to the word "genius", failing to recognize the sarcastic criticism of market analysts. Similarly, common Chinese stock market expressions like "韭菜" (leeks, referring to retail investors being harvested) or "散户" (retail investors) carry implicit sentiment that requires understanding of market culture and context.

Furthermore, traditional approaches typically focus on analyzing isolated text segments, failing to capture the rich context of social media interactions and their temporal dynamics. Consider a news article about a company's positive earnings report with seemingly neutral comments like "又是完美业绩" (another perfect earnings) and "操盘手好厉害" (the trader is impressive). While these comments appear positive in isolation, when analyzed in the context of the company's previous controversial reports and the broader market discussion, they often carry skeptical or negative sentiment. Such contextual nuances, combined with the rapid evolution of Chinese financial terminology and the complex interplay between news and social responses, pose significant challenges for existing sentiment analysis systems. These limitations frequently lead to misinterpretation of market signals, particularly in cases where the true market sentiment emerges not from individual text segments but from the collective discourse patterns and implicit market knowledge.

Current approaches [24] to Chinese financial sentiment analysis suffer from several critical limitations that affect their practical utility. Most existing methods rely heavily on supervised learning with large labeled datasets, making them impractical for Chinese financial markets where such data is scarce. Additionally, these approaches often treat text as isolated segments, failing to capture the complex contextual relationships and cultural nuances inherent in Chinese social media discourse. Traditional models also typically focus on singular aspects of sentiment, such as just the main text content, ignoring the rich ecosystem of interactions and signals available in social media platforms. This narrow focus leads to incomplete and potentially misleading sentiment analysis, particularly when dealing with subtle expressions of market sentiment in Chinese financial discussions.

This paper presents several significant contributions to address the current limitations in Chinese financial sentiment analysis. We propose a novel framework that combines techniques with pre-trained models to effectively leverage limited labeled data while maintaining robust performance. Our context-aware architecture explicitly models the hierarchical nature of social media conversations, enabling better understanding of complex linguistic phenomena in Chinese financial discussions. Furthermore, we introduce a multi-factor fusion approach that considers multiple sentiment signals and their relationships, providing a more comprehensive and accurate sentiment indicator. Through extensive experiments, we demonstrate that our framework significantly improves both sentiment classification accuracy and downstream financial applications, offering a practical solution for real-world Chinese financial sentiment analysis.

Contribution The main contributions of this paper are summarized as follows:

* We propose a framework that effectively utilizes limited labeled data while maintaining high performance in Chinese financial sentiment analysis.
* We develop a novel architecture integrating context-aware modeling and multi-factor fusion mechanisms, capturing complex linguistic phenomena and comprehensive sentiment signals from Chinese financial social media

The remainder of this paper is organized as follows. Section II reviews related work in financial sentiment analysis and its applications. Section III introduces the background of Chinese financial social media analysis and presents our proposed framework, including the context-aware architecture, and multi-factor fusion mechanism. Section IV details our experimental setup, results, and comparative analysis. Finally, Section V concludes the paper and discusses future research directions.

II. Related Work

A. Sentiment Analysis in Financial Domain

Financial sentiment analysis has evolved significantly with advances in machine learning and natural language processing. Traditional machine learning approaches [1], [2], [8], [17] utilized SVM, Naive Bayes, and Random Forest algorithms with manually engineered features such as TF-IDF, bag-of-words, and sentiment lexicons. The emergence of deep learning brought significant improvements through models like CNN for local feature extraction [5], [20], LSTM and GRU for capturing long-term dependencies [12], [18], [22], and attention mechanisms for focusing on relevant information. Transformer-based models, particularly BERT and its variants [3], [10], [26], have achieved state-of-the-art performance by leveraging large-scale pre-training and contextual word embeddings. Recent developments [6], [11], [14], [23] in large language models (LLMs) have further enhanced sentiment analysis capabilities through few-shot learning and better understanding of market context. However, these models primarily focus on English texts, and their direct application to Chinese financial content faces challenges due to linguistic differences and the unique characteristics of Chinese social media discourse. While some studies have attempted to address Chinese-specific issues through specialized pre-training or architectural modifications, they often struggle with limited labeled data and context understanding.

B. Financial Applications of Sentiment Analysis

The application of sentiment analysis in financial markets has demonstrated significant value across various domains. Studies [9], [16], [21] have shown strong correlations between social media sentiment and stock market movements [7], enabling more accurate price prediction and risk assessment. These sentiment indicators have been successfully integrated into trading strategies [19], [25], portfolio management [13], and market surveillance systems [4]. But most existing applications rely on simplified sentiment models that fail to capture the full complexity of market sentiment, particularly in the Chinese market where social media plays a crucial role in investor behavior. The limitations of current approaches highlight the need for more sophisticated sentiment analysis frameworks that can better handle the unique challenges of Chinese financial social media while providing reliable indicators for practical financial applications.

III. Methodology

We propose a comprehensive framework for Chinese financial sentiment analysis that captures contextual information through a hierarchical architecture. Given an input instance $X = \{x_t, x_c, x_{r1}, ..., x_{rn}\}$, where x_t represents the title, x_c the main content, and $\{x_{r1}, ..., x_{rn}\}$ the associated comments, our framework processes these components to generate a set of sentiment predictions $Y = \{y_t, y_c, y_{r1}, ..., y_{r_n}, y_{int}\}$, where y_t, y_c, and y_{ri} represent the sentiment labels for title, content, and each comment respectively, and y_{int} denotes the integrated sentiment that considers the interactions among all components.

A. Input Structure and Preprocessing

Financial social media content consists of diverse components including titles, primary content, and user comments that require structured processing. To effectively analyze these components, we first formalize the input representation and apply systematic preprocessing.

The input is represented as a formatted sequence S:

$$\begin{aligned} S = & [CLS], \\ & t_1, \ldots, t_m, [TITLE], \\ & c_1, \ldots, c_k, [CONTENT], \\ & r_{11}, \ldots, r_{1n}, [COMMENT], \\ & [SEP] \end{aligned} \tag{1}$$

where t_i, c_i, and r_{ij} denote tokens from the title, content, and comments, respectively. Special tokens serve as structural markers to delineate different components and facilitate contextual understanding.

To address the challenge of variable sequence lengths while preserving semantic completeness, dynamic truncation is applied:

$$l_i = \min\left(L_i, \lfloor \alpha_i \cdot L_{\max} \rfloor\right) \tag{2}$$

where L_i represents the original length of the i-th component, $L_{\max}$ is the maximum allowable sequence length, and α_i is a component-specific allocation ratio. We empirically set $\alpha_{\text{title}} = 0.2$, $\alpha_{\text{content}} = 0.5$, and $\alpha_{\text{comments}} = 0.3$ to maintain a balanced representation across components while prioritizing the main content.

The preprocessed input preserves the hierarchical structure of financial social media content while enabling efficient processing by subsequent model components. This formalization ensures consistent handling of the linguistic patterns and domain-specific features present in Chinese financial discussions.

B. Hierarchical Context-Aware Architecture

We implement a hierarchical transformer-based architecture that jointly processes different input components through multiple levels of abstraction. The architecture consists of three integrated components:

1) Hierarchical Encoder: Each input component is first processed through a component-specific encoder layer:

$$h_i^l = \text{LayerNorm}(\text{MHA}(h_i^{l-1}) + h_i^{l-1}) \tag{3}$$
$$h_i^l = \text{LayerNorm}(\text{FFN}(h_i^l) + h_i^l) \tag{4}$$

where MHA is multi-head attention and FFN is a position-wise feed-forward network, h_i^l represents the hidden state of the i-th component at layer l. The hierarchical structure is implemented through component-specific attention masks that control information flow between different parts of the input.

2) Multi-Level Attention and Fusion: Our model integrates both local and global context through a unified attention mechanism:

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V \tag{5}$$

where M is a learnable mask matrix that encodes hierarchical relationships between components. The attention weights are computed as:

$$\begin{cases} \beta_{ij} = \dfrac{\exp(g(h_i, h_j))}{\sum_k \exp(g(h_i, h_k))}, \\ g(h_i, h_j) = v^T \tanh(W_1 h_i + W_2 h_j). \end{cases} \tag{6}$$

where β_{ij} represents the attention weight between components i and j, h_i and h_j are their respective hidden representations, W_1 and W_2 are learnable parameter matrices, and v is a learnable vector. The function $g(h_i, h_j)$ measures the compatibility between two components, while the softmax operation ensures the attention weights sum to 1.

The final representation is obtained through a gated fusion mechanism that combines information from all hierarchical levels:

$$z_f = \sum_{i=1}^{N} g_i \odot h_i \Longrightarrow g_i = \sigma(W_g[h_i; c] + b_g) \tag{7}$$

where $g_i \in \mathbb{R}^d$ are learned gate values for the i-th component, $\odot$ represents element-wise multiplication, N is the number of components, and $h_i \in \mathbb{R}^d$ is the hidden representation of the i-th component. The global context vector $c \in \mathbb{R}^d$ is computed as a weighted sum of all component representations:

$$c = \sum_{i=1}^{N} \alpha_i h_i, \quad \alpha_i = \frac{\exp(w^T h_i)}{\sum_{j=1}^{N} \exp(w^T h_j)} \tag{8}$$

where $w \in \mathbb{R}^d$ is a learnable parameter vector, and α_i represents the importance weight of the i-th component. The concatenation $[h_i; c]$ ensures that each gate value considers both local (h_i) and global (c) context, while $W_g \in \mathbb{R}^{d \times 2d}$ and $b_g \in \mathbb{R}^d$ are learnable parameters. The sigmoid function $\sigma(\cdot)$ ensures gate values are bounded

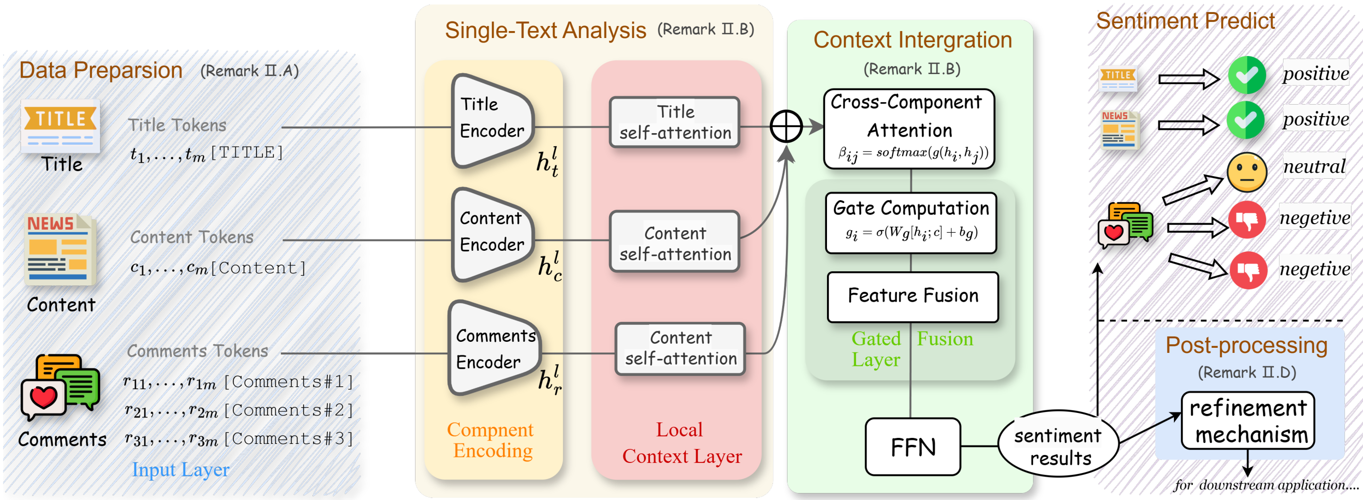

Fig. 2. Overview of the proposed framework for Chinese financial sentiment analysis. The framework consists of four main stages: (1) Data Preparation, which tokenizes title, content, and comments with special tokens; (2) Single-Text Analysis, including component-specific encoding and local self-attention mechanisms for each input type; (3) Context Integration, which uses cross-component attention and gated fusion to combine information from different sources; and (4) Sentiment Prediction, which generates both component-specific sentiments and an integrated sentiment prediction through a feed-forward network (FFN). The framework processes each component independently before integrating contextual information for final sentiment classification.

between 0 and 1, controlling the flow of information from each component to the final representation z_f.

The final sentiment prediction is computed as:

$$\hat{y} = \text{softmax}(W_f z_f + b_f) \tag{9}$$

This integrated architecture allows the model to capture both component-specific features and cross-component relationships while maintaining the hierarchical structure of the input data.

C. Training Strategy

Our model adopts a multi-task learning approach where each component (title, content, comments) contributes to both its local sentiment prediction and the final integrated sentiment. For each component $i \in \{t, c, r\}$, we compute a task-specific cross-entropy loss $\mathcal{L}_i$ for its sentiment classification:

$$\mathcal{L}_i = -\sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k}) \tag{10}$$

where $y_{i,k}$ is the ground truth sentiment label for component i in class k, and $\hat{y}_{i,k}$ is the corresponding predicted probability. The final training objective combines these component-specific losses with an integrated sentiment prediction loss:

$$\mathcal{L}_{total} = \sum_{i \in \{t,c,r\}} \lambda_i \mathcal{L}_i + \lambda_{int} \mathcal{L}_{int} + \gamma \|\theta\|_2^2 \tag{11}$$

where λ_i weights the importance of each component's sentiment task, $\mathcal{L}_{int}$ is the loss for the integrated sentiment prediction, and the last term represents L2 regularization with coefficient γ. We empirically set $\lambda_t = \lambda_c = \lambda_r = 0.3$, $\lambda_{int} = 0.1$, and $\gamma = 0.01$ to balance the contribution of different components while preventing overfitting.

D. Post-processing Sentiment Refinement

Considering the requirements of downstream tasks such as stock movement prediction, we observe that the independent sentiment labels of news and their associated comments often introduce considerable noise in the overall sentiment statistics. To address this challenge, we propose a post-processing refinement mechanism that calibrates news sentiment predictions by leveraging the collective information from comments. Given an initial sentiment prediction y_n for news n and its associated comments $C = \{c_1, c_2, ..., c_m\}$ with sentiment labels $y_c = \{y_1, y_2, ..., y_m\}$, we define two calibration metrics: comment sentiment entropy $H(C)$ and news-comment consistency $\text{Cons}(n, C)$:

$$H(C) = -\sum_{s \in \{-1,0,1\}} p(s) \log_2 p(s) / \log_2 3 \tag{12}$$

$$\text{Cons}(n, C) = \frac{|\{c_i | y_i = y_n, c_i \in C\}|}{|C|} \tag{13}$$

The post-processing refinement function R adjusts the initial sentiment prediction according to these metrics:

$$R(y_n) = \begin{cases} 0 & \text{if } H(C) > \theta_H \\ \text{mode}(y_c) & \text{if } \text{Cons}(n, C) < \theta_C \\ y_n & \text{otherwise} \end{cases} \tag{14}$$

where θ_H and θ_C are predefined thresholds. This calibration process not only helps mitigate the noise from independent sentiment annotations but also provides more reliable sentiment signals for downstream financial applications. Through experimental validation IV-C in stock movement prediction tasks, we demonstrate that this refinement mechanism effectively improves the quality of sentiment features for market analysis.

IV. Experiment

We evaluate our framework through two main experiments: sentiment classification performance on Chinese financial datasets and the effectiveness of post-processing refinement in stock price prediction. Results demonstrate both improved sentiment analysis accuracy and enhanced reliability for downstream financial applications.

A. Experimental Setup

1) Datasets: We evaluate our framework on three Chinese financial datasets that capture different aspects of market sentiment.

TABLE I
Model Comparison

Model	# Parameters	Category	Context Support
RF-SVM [17]	-	Traditional ML	✗
LSTM [12]	15M	Deep Learning	✗
BERT-FIN [3]	340M	Pre-trained LM	✗
BBT-Fin [14]	12B	Large LM	✗
Con-LLAMA	7B	Large LM	✓

1) **EastMoney**[2]: A dataset is collected from EastMoney, China's largest and most influential financial information platform. This dataset contains many posts spanning from January 2023 to December 2023, including stock-specific discussions, market analysis, and user interactions that reflect retail investor sentiment.
2) **FinCUGE** [15]: a comprehensive Chinese stock market sentiment dataset aggregated from multiple social media platforms including Weibo and other financial forums. The multi-platform nature of FinCUGE helps evaluate our model's robustness across different social media contexts.
3) **TCMC**: We also introduce TCMC (Tiered Chinese Market Comments), a new dataset specifically constructed for hierarchical sentiment analysis. TCMC contains 2,000 samples collected from stock discussion boards, where each sample includes a news title, news content, and associated comments. Each component is manually labeled with sentiment polarity by financial domain experts, enabling evaluation of both component-level and integrated sentiment analysis. The inter-annotator agreement achieves a Kappa score of 0.82.

2) Baseline Models: For baseline comparison, we select three representative models: LSTM-CNN, which combines LSTM and CNN layers to capture both sequential patterns and local features in traditional deep learning approaches; BERT-FIN [3], a BERT-based model fine-tuned on Chinese financial texts; and BBT-Fin [14], a state-of-the-art large language model specifically adapted for Chinese financial analysis. Our proposed Con-LLAMA extends the base LLAMA architecture with hierarchical attention mechanisms and context integration modules to better capture the complex relationships in financial social media content.

3) Implementation Details: We implement our framework using PyTorch 2.0 and conduct experiments on a server equipped with dual NVIDIA RTX 4090 GPUs with 24GB memory each, running Ubuntu 22.04.

The model uses a hidden size of 768 dimensions, 12 attention heads, and 12 transformer layers. Training is performed using AdamW optimizer with a learning rate of 2e-5 and a batch size of 32. We apply linear learning rate warmup over the first 10% of steps followed by cosine decay. The maximum sequence length is set to 512 tokens, with component-specific ratios $\alpha_{\text{title}} = 0.2$, $\alpha_{\text{content}} = 0.5$, and $\alpha_{\text{comments}} = 0.3$. Training runs for 10 epochs with early stopping based on validation loss with a patience of 3 epochs. We use gradient clipping with a maximum norm of 1.0 and apply dropout with a rate of 0.1 to prevent overfitting.

B. Sentiment Classification Results

We conduct extensive experiments on the TCMC dataset to evaluate the effectiveness of our proposed framework in hierarchical sentiment analysis. Our Con-LLAMA model achieves superior performance across all average metrics, with an accuracy of 78.1%, precision of 77.3%, recall of 77.9%, and F1-score of 77.6%. This represents substantial improvements over traditional deep learning approaches like LSTM-CNN (F1: 69.9%) and transformer-based models like BERT-FIN (F1: 73.9%). The BBT model, while competitive, achieves an F1-score of 75.9%, demonstrating the advantages of our context-aware architecture.

The performance gains are particularly pronounced in cases involving complex sentiment expressions and multi-component interactions. Con-LLAMA's improvement in precision (3.5% over BBT) and recall (3.8% over BERT-FIN) suggests that our hierarchical attention mechanism effectively captures the nuanced relationships between titles, content, and comments in Chinese financial discussions. The commercial LLMs demonstrate strong performance with context-aware analysis. When provided with contextual information, GPT-3.5-Turbo, Qwen-turbo, and DeepSeek-v3 all show significant improvements in sentiment classification accuracy compared to their single-text counterparts. DeepSeek-v3 achieves the best performance among commercial models, with its context-aware version

[2]Dataset link is here.

TABLE II
PERFORMANCE COMPARISON ON TCMC DATASET (ACC: ACCURACY, P: PRECISION, R: RECALL)

	TITLE				***CONTENT***				***COMMENTS***			
Model	Acc	P	R	F1	Acc	P	R	F1	Acc	P	R	F1
Base Models												
LSTM-CNN	0.700	0.687	0.692	0.685	0.720	0.694	0.711	0.699	0.723	0.700	0.724	0.704
BERT-FIN	0.728	0.729	0.734	0.720	0.745	0.731	0.751	0.734	0.753	0.740	0.746	0.741
BBT-FinT5	0.753	0.747	0.754	0.736	0.761	0.759	0.749	0.749	0.767	0.772	0.764	0.764
Con-LLAMA (Ours)	**0.777**	**0.764**	**0.767**	**0.756**	**0.783**	**0.772**	**0.766**	**0.774**	**0.789**	**0.790**	**0.785**	**0.777**
Context Enhancement Analysis on Commercial LLMs[1]												
GPT-3.5-Turbo (single)	0.748	0.753	0.756	0.745	0.753	0.756	0.753	0.759	0.743	0.752	0.755	0.758
GPT-3.5-Turbo (**context**)	0.792	0.787	0.785	0.771	0.805	0.784	0.785	0.792	0.797	0.791	0.794	0.783
Qwen-turbo (single)	0.762	0.758	0.748	0.755	0.762	0.754	0.759	0.767	0.770	0.770	0.764	0.774
Qwen-turbo (**context**)	0.793	0.786	0.791	0.784	0.805	0.783	0.798	0.789	0.803	0.791	0.797	0.784
DeepSeek-v3 (single)	0.769	0.757	0.758	0.749	0.767	0.748	0.751	0.745	0.761	0.752	0.758	0.758
DeepSeek-v3 (**context**)	0.791	0.782	0.790	0.777	0.815	0.786	0.790	0.784	0.808	0.800	0.803	0.797

[1] "single" means only inputting the target content, while "context" means providing context (through prompt engineering)

reaching accuracy scores of 0.815 on content analysis and 0.808 on comments analysis. The performance gains between single and context versions (an average increase of 3-4 percentage points across all metrics) validate the importance of considering contextual information in sentiment analysis for Chinese financial texts. Details about prompt engineering seen in Appendix V.

Fig. 3. Performance comparison of different sentiment analysis models on EastMoney and FinCUGE datasets. Our proposed Con-LLAMA model achieves superior performance across F1-score, accuracy, and recall metrics compared to baseline models including RF-SVM, LSTM-CNN, BERT-FIN, and BBT-FinTS.

When evaluating on EastMoney and FinCUGE datasets, which lack the hierarchical structure present in TCMC, our Con-LLAMA model demonstrates comparable performance to the standard LLAMA-based approaches. On EastMoney, Con-LLAMA achieves an F1-score of 76.2%, showing marginal difference from BBT's 75.9% and single LLAMA's 76.0%. Similarly, on FinCUGE, our model achieves an F1-score of 74.7%, closely matching the performance of single LLAMA (74.5%). These results are expected as these datasets do not contain the structured format of news titles, content, and comments that our model is specifically designed to leverage. However, the comparable performance demonstrates that our context-aware architecture maintains robust sentiment analysis capabilities even when operating on simpler, unstructured text formats, without introducing additional overhead or performance degradation.

C. Downstream Application

To validate the practical utility of our sentiment refinement mechanism, we conduct a case study on a representative stock from the CSI 300 index over an 80-day trading period. We implement a sliding window prediction model that utilizes the previous 50 days' data to predict the stock price on day 51, combining traditional technical indicators with sentiment signals. Figure 3 illustrates the prediction results comparing different sentiment approaches.

The post-processing sentiment refinement demonstrates substantial improvements in prediction accuracy compared to the baseline sentiment analysis without refinement. Using sentiment signals with post-processing refinement achieves a Mean Absolute Error (MAE) of 1.1245 and Mean Square Error (MSE) of 1.9037, significantly outperforming the standard sentiment approach (MAE: 1.4044, MSE: 2.9964). As shown in Figure 3, the refined sentiment signals (red dashed line) track the ground truth (solid black line) more closely than the normal sentiment approach (blue dashed line), particularly during periods of high price volatility (e.g., trading days 70-90). This improvement suggests that our comment-aware refinement mechanism effectively reduces noise in sentiment signals and provides more reliable indicators for stock price prediction. Future work will extend this analysis to a broader range of stocks to validate the generalizability of our approach.

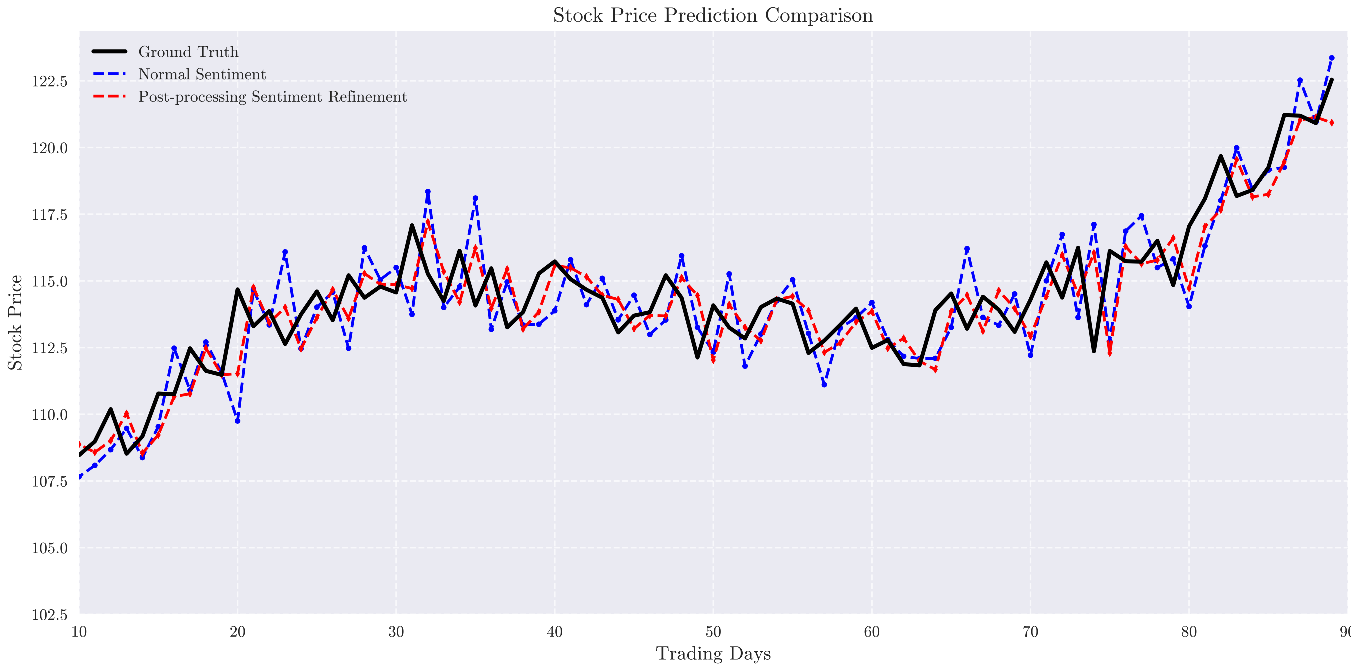

Fig. 4. Comparison of stock price prediction performance between different sentiment approaches over 80 trading days. The post-processing sentiment refinement (red dashed line) shows better alignment with ground truth (black solid line) compared to normal sentiment analysis (blue dashed line), particularly during high volatility periods (days 70-90).

D. Ablation Studies

To understand the contribution of each component in our proposed framework, we conduct comprehensive ablation studies. We systematically remove key components: (1) hierarchical attention mechanism, (2) context integration module. Removing the hierarchical attention mechanism leads to a 2.8% drop in sentiment classification F1-score (from 77.6% to 74.8%) and a corresponding increase in stock prediction MSE by 0.0023. The absence of context integration results in a 2.1% decrease in F1-score, particularly affecting the model's ability to interpret complex, multi-part financial discussions. Without the multi-factor fusion layer, performance drops by 1.9%, indicating its importance in combining signals from different components.

TABLE III
Ablation Analysis

Key Component	F1-Score	Accuracy	MSE
Original Model	0.728	0.731	0.221
+ Attention mechanism	0.748	0.752	0.198
+ Context integration	0.757	0.762	0.186
Con-LLAMA **(OURS)**	**0.786**	**0.781**	**0.0142**

These results demonstrate that each component plays a crucial role in the overall framework, with the hierarchical attention mechanism contributing most significantly to the model's performance. The ablation studies also reveal that the components work synergistically - while individual components show moderate performance impacts, their combination enables the model to effectively capture the complex relationships in Chinese financial social media content.

V. Conclusion

This paper presents a novel framework for Chinese financial sentiment analysis that effectively addresses the challenges of analyzing complex social media content in financial markets. Our approach combines hierarchical attention mechanisms with context-aware architecture to capture the nuanced relationships between different components of financial discussions. Through extensive experiments on multiple Chinese financial datasets, we demonstrate that our model achieves superior performance in both sentiment classification and downstream financial applications.

While our results show promising advances in Chinese financial sentiment analysis, several directions for future research remain. These include exploring the integration of market microstructure data, extending the framework to handle real-time streaming data, and investigating the model's adaptability to different financial market contexts. We believe our work provides a solid foundation for future developments in financial sentiment analysis and its applications in market prediction.

References

[1] Yassine Al Amrani, Mohamed Lazaar, and Kamal Eddine El Kadiri. Random forest and support vector machine based

hybrid approach to sentiment analysis. *Procedia Computer Science*, 127:511–520, 2018.

[2] Ghaith Abdulsattar A Jabbar Alkubaisi, Siti Sakira Kamaruddin, and Husniza Husni. Stock market classification model using sentiment analysis on twitter based on hybrid naive bayes classifiers. *Comput. Inf. Sci.*, 11(1):52–64, 2018.

[3] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models, 2019.

[4] Michael Breen, Dermot Hodson, and Manuela Moschella. Incoherence in regime complexes: A sentiment analysis of eu-imf surveillance. *JCMS: Journal of common market studies*, 58(2):419–437, 2020.

[5] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen. Financial time-series data analysis using deep convolutional neural networks. In *2016 7th International conference on cloud computing and big data (CCBD)*, pages 87–92. IEEE, 2016.

[6] Xiang Deng, Vasilisa Bashlovkina, Feng Han, Simon Baumgartner, and Michael Bendersky. What do llms know about financial markets? a case study on reddit market sentiment analysis. In *Companion Proceedings of the ACM Web Conference 2023*, pages 107–110, 2023.

[7] Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria. Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 56(9):1–42, 2024.

[8] Ankur Goel, Jyoti Gautam, and Sitesh Kumar. Real time sentiment analysis of tweets using naive bayes. In *2016 2nd international conference on next generation computing technologies (NGCT)*, pages 257–261. IEEE, 2016.

[9] Xinyi Guo and Jinfeng Li. A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 472–477. IEEE, 2019.

[10] Joshua Zoen Git Hiew, Xin Huang, Hao Mou, Duan Li, Qi Wu, and Yabo Xu. Bert-based financial sentiment index and lstm-based stock return predictability. *arXiv preprint arXiv:1906.09024*, 2019.

[11] Giorgos Iacovides, Thanos Konstantinidis, Mingxue Xu, and Danilo Mandic. Finllama: Llm-based financial sentiment analysis for algorithmic trading. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 134–141, 2024.

[12] Ching-Ru Ko and Hsien-Tsung Chang. Lstm-based sentiment analysis for stock price forecast. *PeerJ Computer Science*, 7:e408, 2021.

[13] Yuchen Liu, Daniil Mikriukov, Owen Christopher Tjahyadi, Gangmin Li, Terry R Payne, Yong Yue, Kamran Siddique, and Ka Lok Man. Revolutionising financial portfolio management: The non-stationary transformer's fusion of macroeconomic indicators and sentiment analysis in a deep reinforcement learning framework. *Applied Sciences*, 14(1):274, 2023.

[14] Dakuan Lu, Hengkui Wu, Jiaqing Liang, Yipei Xu, Qianyu He, Yipeng Geng, Mengkun Han, Yingsi Xin, and Yanghua Xiao. Bbt-fin: Comprehensive construction of chinese financial domain pre-trained language model, corpus and benchmark, 2023.

[15] Maciel. Fincuge-instruction, 2023. Dataset; Accessed: 2025-1-11, License: Apache-2.0.

[16] Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. Stock price prediction using news sentiment analysis. In *2019 IEEE fifth international conference on big data computing service and applications (BigDataService)*, pages 205–208. IEEE, 2019.

[17] Faithful Chiagoziem Onwuegbuche, Joseph Muliaro Wafula, Joseph Kyalo Mung'atu, et al. Support vector machine for sentiment analysis of nigerian banks financial tweets. *Journal of Data Analysis and Information Processing*, 7(04):153, 2019.

[18] Aakanksha Sharaff, Tushin Roy Chowdhury, and Sakshi Bhandarkar. Lstm based sentiment analysis of financial news. *SN Computer Science*, 4(5):584, 2023.

[19] Baljinder Singh, Mandeep Kaur, and Gurbinder Singh Brar. Enhancing financial trading strategies: A comprehensive sentiment analysis pipeline. In *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)*, volume 1, pages 1–7. IEEE, 2024.

[20] Sahar Sohangir, Dingding Wang, Anna Pomeranets, and Taghi M Khoshgoftaar. Big data: Deep learning for financial sentiment analysis. *Journal of Big Data*, 5(1):1–25, 2018.

[21] Chuan-Ju Wang, Ming-Feng Tsai, Tse Liu, and Chin-Ting Chang. Financial sentiment analysis for risk prediction. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 802–808, 2013.

[22] You Wu, Mengfang Sun, Hongye Zheng, Jinxin Hu, Yingbin Liang, and Zhenghao Lin. Integrative analysis of financial market sentiment using cnn and gru for risk prediction and alert systems. In *2024 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, pages 410–415. IEEE, 2024.

[23] Frank Xing. Designing heterogeneous llm agents for financial sentiment analysis. *ACM Transactions on Management Information Systems*, 2024.

[24] Chenrui Zhang. Research on the correlation between text emotion mining and stock market based on deep learning. *arXiv preprint arXiv:2205.06675*, 2022.

[25] Wenbin Zhang and Steven Skiena. Trading strategies to exploit blog and news sentiment. In *Proceedings of the international AAAI conference on web and social media*, pages 375–378, 2010.

[26] Lingyun Zhao, Lin Li, Xinhao Zheng, and Jianwei Zhang. A bert based sentiment analysis and key entity detection approach for online financial texts. In *2021 IEEE 24th International conference on computer supported cooperative work in design (CSCWD)*, pages 1233–1238. IEEE, 2021.

Appendix A: Prompt Engineering Design

A.1 Prompt Templates

For commercial LLMs (GPT-3.5-Turbo, Qwen-turbo, and DeepSeek-v3), we designed two types of prompts:

Single-text Analysis Prompt:

```
You are a professional financial analyst.
Please analyze the sentiment of the following
Chinese financial text and classify it as
positive (1), negative (-1), or neutral
(0).Only output the sentiment label.

Text: [input_text]
```

Context-aware Analysis Prompt:

```
You are a professional financial analyst. Please
analyze the sentiment of the target content considering
its related context. The sentiment should be
classified  as positive (1), negative (-1), or neutral
(0). Only  output the sentiment label.

Title: [title_text]
Content: [content_text]
Related Comments:[comments_text]
Target Content: [target_text]
```

Each prompt template is designed to enforce structured and consistent sentiment analysis across different commercial LLMs while maintaining clear boundary conditions for sentiment classification.