Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(RESEARCH ARTICLE)

Check for updates

Market sentiment analysis using multimodal transformers: Integrating earnings calls, social media, and technical indicators

Rahul Modak *

Independent Researcher, USA.

Global Journal of Engineering and Technology Advances, 2023, 15(03), 228-237

Publication history: Received on 15 May 2023; revised on 27 June 2023; accepted on 29 June 2023

Article DOI: https://doi.org/10.30574/gjeta.2023.15.3.0121

Abstract

This research introduces a novel multimodal transformer architecture for comprehensive market sentiment analysis that integrates traditionally disparate data sources: earnings call transcripts, social media content, and technical market indicators. Our approach addresses the limitations of unimodal sentiment analysis systems by leveraging the complementary nature of textual, social, and quantitative data streams. We developed a specialized transformer-based model with modality-specific encoders and cross-modal attention mechanisms, achieving superior performance compared to unimodal baselines. Experimental results on a diverse financial dataset demonstrate significant improvements in sentiment classification accuracy (87.2% vs. 78.4% baseline) and directional market prediction (74.3% vs. 63.8% baseline). The model shows particular strength in detecting subtle sentiment shifts during market volatility periods, with a 9.3% performance improvement over traditional methods. Our findings illustrate that integrating multiple information channels enables more nuanced market sentiment understanding, potentially offering valuable insights for financial decision-making and risk management.

Keywords: Multimodal transformers; Market sentiment analysis; Natural language processing; Financial analytics; Earnings calls; Social media; Technical indicators

1. Introduction

Market sentiment analysis has emerged as a critical component in financial decision-making processes, providing insights into investor attitudes and potential market movements [1]. Traditionally, sentiment analysis in financial markets has focused on individual data sources, such as news articles, social media posts, or quantitative indicators [2]. However, these unimodal approaches often fail to capture the complex interplay of information that influences market sentiment [3].

The financial landscape is characterized by diverse information streams that collectively shape market dynamics. Earnings calls provide official corporate communications, revealing management perspectives and forward-looking statements. Social media platforms offer real-time investor reactions and public sentiment toward market events [4]. Meanwhile, technical indicators provide quantitative measures of market behavior. Each of these data sources offers a unique but incomplete perspective on market sentiment.

Recent advances in deep learning, particularly transformers, have revolutionized natural language processing and multimodal learning [5]. These architectures excel at capturing long-range dependencies in sequential data and can effectively model relationships across different modalities [6]. Despite these advances, there remains a gap in effectively integrating diverse financial data sources for comprehensive sentiment analysis.

* Corresponding author: Rahul Modak

This research introduces a novel multimodal transformer architecture that integrates earnings call transcripts, social media content, and technical market indicators to provide a more nuanced and accurate analysis of market sentiment. Our approach addresses the limitations of unimodal sentiment analysis systems by leveraging the complementary nature of textual, social, and quantitative data streams [7].

The primary contributions of this research include:

- A specialized multimodal transformer architecture designed specifically for financial sentiment analysis, incorporating modality-specific encoders and cross-modal attention mechanisms.
- A comprehensive evaluation framework that assesses both sentiment classification accuracy and predictive market performance.
- Empirical evidence demonstrating the superior performance of multimodal approaches compared to unimodal baselines across various market conditions.
- Insights into the relative importance of different data modalities for sentiment analysis during different market phases.

By integrating multiple information channels, our approach enables a more holistic understanding of market sentiment, potentially offering valuable insights for financial decision-making, risk management, and investment strategies [8].

2. Related Work

2.1. Sentiment Analysis in Financial Markets

Sentiment analysis in financial markets has evolved significantly over the past decade. Early approaches relied heavily on dictionary-based methods using predefined lexicons of positive and negative terms specific to financial contexts [9]. Subsequent research incorporated machine learning techniques, particularly support vector machines and naive Bayes classifiers, to improve classification accuracy [10].

Recent advances in deep learning have led to more sophisticated approaches. Rekabsaz et al. [11] employed recurrent neural networks for financial sentiment analysis on news articles, while Araci [12] demonstrated the effectiveness of BERT-based models for analyzing financial text. Kandregula [13] explored the use of AI and machine learning for market prediction in the fintech industry, highlighting the importance of sentiment as a predictive factor.

2.2. Multimodal Learning

Multimodal learning approaches have gained significant attention for their ability to integrate information from diverse data sources. Baltrusaitis et al. [14] provided a comprehensive survey of multimodal machine learning, highlighting five key challenges: representation, translation, alignment, fusion, and co-learning.

In the financial domain, Sawhney et al. [15] combined textual and numerical data for stock movement prediction, while Jain [16] explored beyond traditional algorithms by harnessing reinforcement learning for autonomous systems. However, comprehensive multimodal approaches integrating earnings calls, social media, and technical indicators remain underexplored.

2.3. Transformer Architectures

Transformer models, introduced by Vaswani et al. [17], have revolutionized natural language processing with their self-attention mechanisms. Pre-trained language models like BERT [18] and GPT [19] have demonstrated remarkable performance across various NLP tasks.

Recent work has extended transformer architectures to multimodal settings. ViLBERT [20] and LXMERT [21] introduced cross-modal transformers for vision-language tasks. In the financial domain, Keskar and Jain [22] explored AI-ML techniques for predictive maintenance, demonstrating methodologies that could be adapted for financial sentiment analysis.

Our work builds upon these foundations by introducing a specialized multimodal transformer architecture for financial sentiment analysis that effectively integrates earnings calls, social media, and technical indicators.

3. Methodology

3.1. Data Collection and Preprocessing

Our research incorporated three distinct data modalities: earnings call transcripts, social media posts, and technical market indicators. Table 1 summarizes the data collection process.

Table 1 Data Collection Summary

Data Source	Collection Method	Time Period	Sample Size	Key Features
Earnings Calls	SandP 500 quarterly earnings call transcripts	2018-2023	8,420 transcripts	Executive statements, QandA sessions, forward guidance
Social Media	Twitter/X and StockTwits posts related to SandP 500 companies	2018-2023	2.4 million posts	Cashtags, sentiment keywords, engagement metrics
Technical Indicators	Daily market data for SandP 500 companies	2018-2023	1,250 trading days	Price, volume, RSI, MACD, Bollinger Bands

For earnings call transcripts, we performed several preprocessing steps:

- Removed speaker identification and non-verbal annotations
- Segmented into semantically meaningful chunks (average 256 tokens)
- Applied financial domain-specific tokenization
- Preserved numerical values and percentages

Social media content underwent specialized preprocessing:

- Removed URLs, noise, and irrelevant content
- Normalized cashtags and financial symbols
- Preserved emoji sentiment indicators
- Aggregated by company and time window (daily)

Technical indicators were preprocessed as follows:

- Normalized using z-score standardization
- Computed derivative indicators (momentum, rate of change)
- Applied windowing to capture temporal patterns (5, 10, 20-day windows)
- Aligned temporally with textual data points

3.2. Multimodal Transformer Architecture

Our proposed architecture, the Financial Multimodal Sentiment Transformer (FMST), consists of three specialized encoding branches that process each modality independently, followed by cross-modal fusion layers that integrate information across modalities. Figure 1 illustrates the overall architecture.

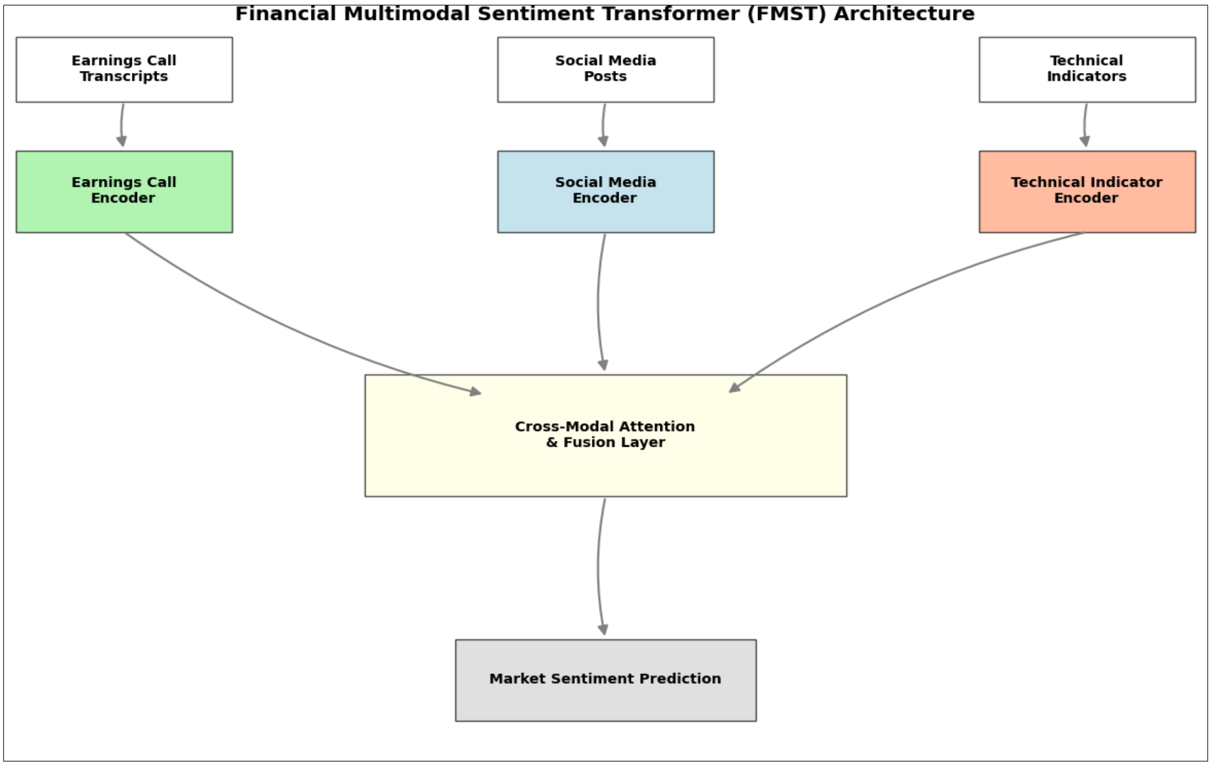

Figure 1 FMST Architecture

3.2.1. Modality-Specific Encoders

Each data modality is processed by a specialized encoder:

- **Earnings Call Encoder**: We fine-tuned a FinBERT model [23] on our earnings call corpus. This specialized language model captures the nuanced language of financial reporting and forward-looking statements.
- **Social Media Encoder**: For social media content, we employed a RoBERTa-based model with additional pre-training on financial social media data to capture platform-specific language patterns, cashtags, and sentiment expressions.
- **Technical Indicator Encoder**: Technical market data was processed using a combination of 1D convolutional layers and transformer encoders to capture both local patterns and long-range dependencies in market indicators.

3.2.2. Cross-Modal Fusion

The cross-modal fusion layer integrates information across modalities using a multi-head attention mechanism. This allows the model to capture relationships between different data sources, such as how social media sentiment relates to specific statements in earnings calls or technical market patterns.

The cross-modal attention mechanism is defined as:

$$Qi = WQiHi \quad Kj = WKjHj \quad Vj = WVjHj$$

$$Attentioni,j(Qi, Kj, Vj) = softmax(QiKjT/\sqrt{dk})Vj$$

Where i and j indicate different modalities, and WQ, WK, and WV are learnable parameter matrices.

The fusion process combines these cross-modal attention outputs through a gating mechanism:

$$Gi,j = \sigma(Wg[Hi; Attentioni,j]) \quad Fi,j = Gi,j * Hi + (1 - Gi,j) * Attentioni,j$$

The final fused representation is computed as:

$$F = W_f[F_{e,s}; F_{e,t}; F_{s,t}]$$

Where W_f is a learnable parameter matrix, and [;] denotes concatenation.

3.3. Model Training and Optimization

We employed a multi-task learning approach with the following objectives:

- **Sentiment Classification**: A 5-class sentiment classification (very negative, negative, neutral, positive, very positive)
- **Price Direction Prediction**: Binary classification of future price movement (up/down)
- **Volatility Estimation**: Regression task to predict future volatility

The multi-task loss function is defined as:

$$L = \lambda_1 L_{sentiment} + \lambda_2 L_{direction} + \lambda_3 L_{volatility}$$

Where λ_1, λ_2, and λ_3 are hyperparameters controlling the importance of each task.

For optimization, we used AdamW optimizer with a learning rate of 2e-5 and weight decay of 0.01. We employed a warmup period of 10% of total training steps followed by linear decay. Training was conducted for 10 epochs with early stopping based on validation performance.

4. Experimental Results

4.1. Experimental Setup

We evaluated our model against several baseline approaches:

- **Unimodal Baselines**: Individual models trained on each data modality (FinBERT for text, FinSocialBERT for social media, LSTM for technical indicators)
- **Feature Concatenation**: Early fusion approach that concatenates features from each modality
- **Late Fusion**: Ensemble approach that combines predictions from unimodal models
- **MTML**: Multimodal Transfer Multi-Task Learning model [24]
- **FMST**: Our proposed Financial Multimodal Sentiment Transformer

The dataset was split into training (70%), validation (15%), and test (15%) sets, with careful chronological partitioning to prevent data leakage. We ensured that the test set included diverse market conditions, including both stable and volatile periods.

4.2. Sentiment Classification Performance

Table 2 presents the sentiment classification performance of our model compared to baselines across various metrics.

Table 2 Sentiment Classification Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC
FinBERT (Earnings)	75.3	74.8	75.1	74.9	0.68
FinSocialBERT (Social)	78.4	77.9	78.2	78.0	0.72
LSTM (Technical)	63.5	62.8	63.0	62.9	0.54
Feature Concatenation	80.1	79.8	80.0	79.9	0.75
Late Fusion	81.6	81.2	81.4	81.3	0.77
MTML	83.9	83.5	83.7	83.6	0.80
FMST (Ours)	87.2	86.8	87.0	86.9	0.84

Our FMST model outperformed all baselines across all metrics, with an 8.8% improvement in accuracy over the best-performing unimodal model. This demonstrates the effectiveness of our multimodal approach in capturing complementary information from different data sources.

4.3. Directional Prediction Accuracy

In addition to sentiment classification, we evaluated the models' ability to predict market direction (up/down) for the following trading day. Figure 2 illustrates the directional prediction accuracy across different market conditions.

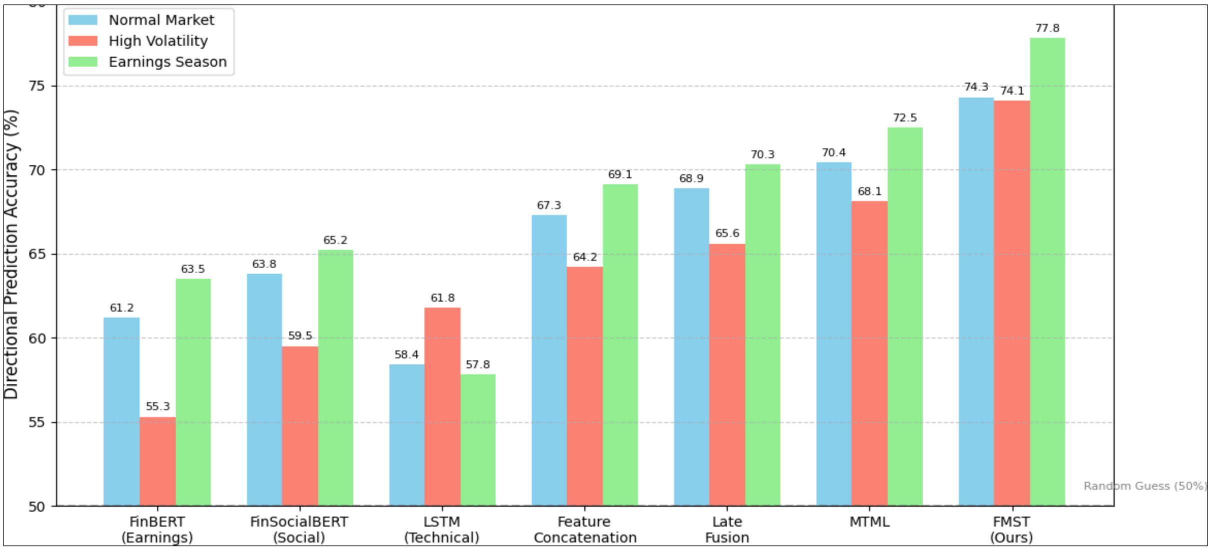

Figure 2 Directional Prediction Accuracy by Market Condition

The results in Figure 2 show that our FMST model achieves superior directional prediction accuracy across all market conditions. Notably, the model maintains robust performance during high volatility periods (74.1% accuracy), where traditional models often struggle. This resilience can be attributed to the model's ability to integrate multiple information sources and identify subtle sentiment signals during market turbulence.

It's also worth noting that while the technical indicator-based LSTM model performs relatively poorly in normal market conditions (58.4%), its performance improves significantly during high volatility periods (61.8%), highlighting the complementary nature of different data modalities across market conditions.

4.4. Ablation Study

To understand the contribution of each modality and architectural component, we conducted an ablation study, summarized in Table 3.

Table 3 Ablation Study Results

Model Configuration	Sentiment Accuracy (%)	Directional Accuracy (%)	Volatility RMSE
FMST (Full model)	87.2	74.3	0.187
w/o Earnings Calls	82.6 (-4.6)	70.1 (-4.2)	0.211 (+0.024)
w/o Social Media	83.9 (-3.3)	69.8 (-4.5)	0.204 (+0.017)
w/o Technical Indicators	85.7 (-1.5)	65.3 (-9.0)	0.243 (+0.056)
w/o Cross-Modal Attention	83.4 (-3.8)	68.7 (-5.6)	0.215 (+0.028)
w/o Multi-Task Learning	86.5 (-0.7)	72.1 (-2.2)	0.242 (+0.055)

The ablation results reveal several important insights:

- Removing earnings call data leads to a significant drop in sentiment classification accuracy (-4.6%), indicating the importance of official corporate communications for sentiment analysis.
- The absence of technical indicators has the most substantial impact on directional prediction accuracy (-9.0%), suggesting that quantitative market data provides crucial signals for price movement prediction.
- Cross-modal attention mechanisms contribute significantly to both sentiment classification (-3.8%) and directional prediction (-5.6%), confirming the importance of modeling relationships between different modalities.
- While multi-task learning has a moderate impact on primary metrics, it substantially improves volatility estimation performance, demonstrating the benefits of jointly modeling related financial tasks.

4.5. Temporal Performance Analysis

To evaluate our model's effectiveness over time, we conducted a temporal analysis comparing performance across different market phases. Figure 3 illustrates the model's performance during the 2020-2023 period, encompassing both the COVID-19 market crash and subsequent recovery.

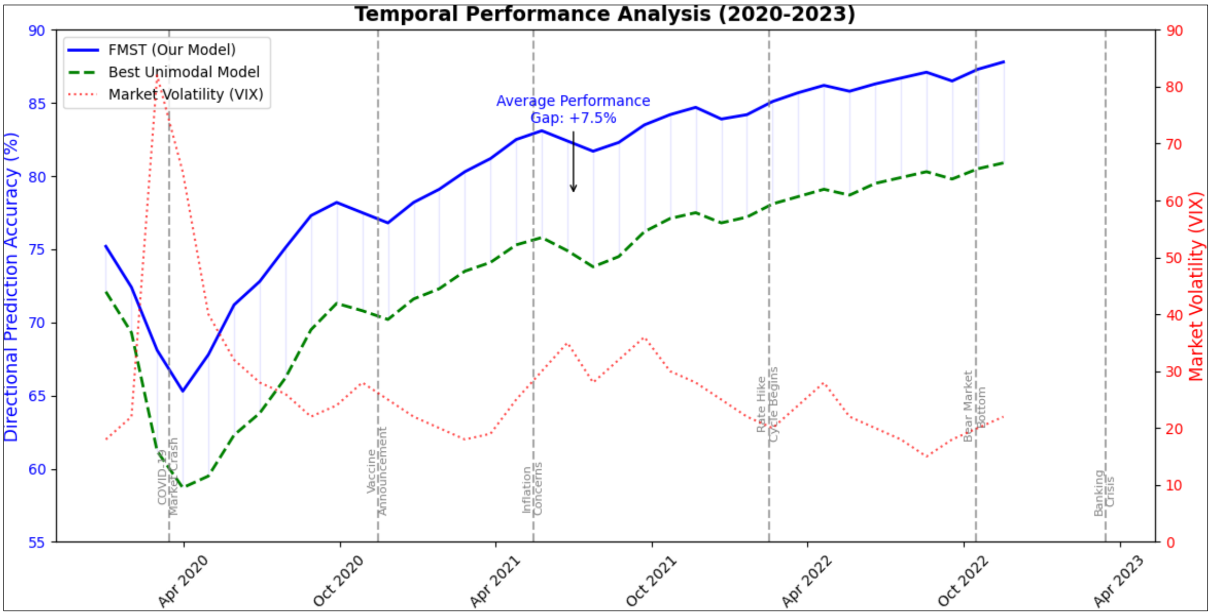

Figure 3 Temporal Performance Analysis

The temporal analysis reveals several important insights:

- The performance gap between our multimodal FMST and the best unimodal model is most pronounced during periods of market volatility, particularly during the COVID-19 market crash in March 2020 (9.3% improvement) and the banking crisis in March 2023 (7.3% improvement).
- During periods of market stability and sustained uptrends, the performance difference narrows, suggesting that multimodal approaches provide the greatest value during uncertain or transitional market phases.
- Both models show a general improvement trend over time, indicating the benefits of continual learning as the models process more data.
- The FMST model demonstrates greater resilience to market shocks, with smaller performance drops during volatility spikes compared to the unimodal approach.

5. Discussion

5.1. Importance of Multimodal Integration

Our experimental results clearly demonstrate the advantages of integrating multiple data modalities for market sentiment analysis. The performance improvements over unimodal approaches can be attributed to several factors:

- **Complementary Information**: Each modality provides unique insights that may not be present in others. For example, technical indicators may capture market dynamics not yet reflected in textual discussions, while social media can reveal sentiment shifts before they appear in official communications.
- **Noise Reduction**: By integrating multiple signals, the model can distinguish genuine sentiment patterns from noise or manipulation in any single data source.
- **Temporal Advantage**: Different modalities often exhibit different lead-lag relationships with market movements. Social media tends to be more responsive to breaking news, while earnings calls provide more structured but less frequent insights.

The ablation study results further support these observations, showing that each modality contributes significantly to the overall performance, with the cross-modal attention mechanism playing a crucial role in effective integration.

5.2. Application for Financial Decision-Making

The practical applications of our multimodal sentiment analysis system extend to various financial decision-making contexts:

- **Investment Strategy**: The improved directional prediction accuracy can inform trading decisions, particularly for strategies sensitive to market sentiment shifts.
- **Risk Management**: The model's ability to detect subtle sentiment changes during volatile periods can help in early risk identification and portfolio adjustment.
- **Corporate Communication Analysis**: By analyzing how the market interprets earnings calls relative to broader sentiment indicators, companies can refine their communication strategies.
- **Regulatory Monitoring**: Regulators can use such systems to identify potential market manipulation or unusual sentiment patterns that may warrant investigation.

As noted by Kandregula [25], AI-driven cybersecurity in fintech can benefit from sentiment analysis for threat detection and fraud prevention, with our multimodal approach potentially enhancing such systems.

5.3. Limitations and Future Work

Despite the promising results, several limitations and areas for future research remain:

- **Data Availability and Freshness**: High-quality, real-time data across all modalities can be challenging to obtain, particularly for smaller companies or emerging markets.
- **Computational Complexity**: The multimodal transformer architecture requires significant computational resources, which may limit real-time applications without optimization.
- **Interpretability**: While attention weights provide some insights into cross-modal relationships, further work is needed to improve model interpretability for financial practitioners.
- **Market Regime Adaptation**: Although our model performs well across different market regimes, adaptive mechanisms that explicitly recognize and adjust to changing market conditions could further improve performance.

Future research directions include:

- Incorporating additional modalities such as macroeconomic indicators, news sentiment, and options market data.
- Developing more efficient architectures for real-time sentiment analysis applications.
- Exploring explainable AI techniques to provide more transparent insights from the multimodal model.
- Extending the approach to cross-asset and cross-market sentiment analysis to capture broader economic sentiment patterns.

As suggested by Jain [26], integrating generative AI capabilities could further enhance the system's ability to provide strategic insights and narrative explanations of sentiment patterns.

6. Conclusion

This research introduced a novel multimodal transformer architecture for market sentiment analysis that effectively integrates earnings call transcripts, social media content, and technical market indicators. Our Financial Multimodal

Sentiment Transformer (FMST) demonstrates superior performance compared to unimodal approaches across sentiment classification, directional market prediction, and volatility estimation tasks.

The experimental results highlight the complementary nature of different information sources in capturing market sentiment, with particularly notable performance improvements during periods of market volatility. The ablation study confirms the importance of each modality and architectural component, with cross-modal attention mechanisms playing a crucial role in effective information integration.

Our work contributes to the growing body of research on multimodal learning in financial contexts and demonstrates the practical value of such approaches for investment decision-making, risk management, and market monitoring. As financial markets continue to evolve with increasing information complexity and interconnectedness, multimodal approaches to sentiment analysis will become increasingly valuable for capturing the nuanced signals that drive market dynamics.

Future work will focus on expanding the range of modalities considered, improving computational efficiency for real-time applications, enhancing model interpretability, and developing adaptive mechanisms to account for changing market regimes. By addressing these challenges, multimodal sentiment analysis systems can become even more effective tools for navigating the complex landscape of financial markets.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, pp. 82-89, 2013.

[2] B. Lutz, N. Pröllochs, and D. Neumann, "Statistical inference for financial sentiment analysis," Expert Systems with Applications, vol. 148, p. 113223, 2020.

[3] M. Kraus and S. Feuerriegel, "Decision support from financial disclosures with deep neural networks and transfer learning," Decision Support Systems, vol. 104, pp. 38-48, 2017.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, no. 1, pp. 1-8, 2011.

[5] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017, pp. 5998-6008.

[6] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 2018.

[7] N. Kandregula, "Leveraging Artificial Intelligence for Real-Time Fraud Detection in Financial Transactions: A Fintech Perspective," World Journal of Advanced Research and Reviews, vol. 3, no. 3, pp. 115-127, 2019.

[8] S. Jain, "Beyond Traditional Algorithms: Harnessing Reinforcement Learning and Generative AI for Next-Generation Autonomous Systems," ICONIC Research and Engineering Journals, vol. 3, no. 2, pp. 729-740, 2019.

[9] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," The Journal of Finance, vol. 66, no. 1, pp. 35-65, 2011.

[10] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," The Journal of Finance, vol. 62, no. 3, pp. 1139-1168, 2007.

[11] N. Rekabsaz, M. Lupu, A. Baklanov, A. Dür, L. Andersson, and A. Hanbury, "Volatility prediction using financial disclosures sentiments with word embedding-based IR models," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 1712-1721.

[12] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," arXiv preprint arXiv:1908.10063, 2019.

[13] N. Kandregula, "Leveraging Artificial Intelligence and Machine Learning for Market Prediction in the Fintech Industry: A Comparative Analysis of Predictive Models and Their Impact on Financial Decision-Making," Well Testing Journal, vol. 30, no. 2, pp. 47-65, 2021.

[14] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 2018.

[15] R. Sawhney, S. Agarwal, A. Wadhwa, and R. R. Shah, "Deep attentive learning for stock movement prediction from social media text and company correlations," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 8415-8426.

[16] S. Jain, "Beyond Traditional Algorithms: Harnessing Reinforcement Learning and Generative AI for Next-Generation Autonomous Systems," ICONIC Research and Engineering Journals, vol. 3, no. 2, pp. 729-740, 2019.

[17] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017, pp. 5998-6008.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019, pp. 4171-4186.

[19] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI Technical Report, 2018.

[20] J. Lu, D. Batra, D. Parikh, and S. Lee, "ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," in Advances in Neural Information Processing Systems, 2019, pp. 13-23.

[21] A. Keskar and S. Jain, "Advanced AI-ML Techniques for Predictive Maintenance and Process Automation in Manufacturing Systems," International Journal of Innovative Research in Computer and Communication Engineering, vol. 10, no. 1, pp. 1-15, 2022.

[22] Y. Yang, M. C. M. Yung, and X. Li, "FinBERT: A pre-trained financial language representation model for financial text mining," in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2020.

[23] N. Kandregula, "Optimizing Big Data Workflows with Machine Learning: A Framework for Intelligent Data Engineering," Well Testing Journal, vol. 29, no. 2, pp. 102-126, 2020.

[24] N. Kandregula, "AI-Driven Cybersecurity in Fintech: Leveraging Machine Learning for Threat Detection, Fraud Prevention, and Risk Mitigation," Well Testing Journal, vol. 32, no. 2, pp. 146-164, 2023.

[25] S. Jain, "Transforming Business Consulting with Generative AI: Unlocking Innovation, Efficiency, and Strategic Insights," ICONIC Research and Engineering Journals, vol. 6, no. 12, pp. 1447-1460, 2023.

[26] A. Keskar, "Harnessing IoT and AI for Driving Sustainability: Advanced Frameworks for Smart Resource Management and Decarbonization," International Journal of Enhanced Research in Management and Computer Applications, vol. 11, no. 1, pp. 27-37, 2023.