



Multimodal Deep Learning for Financial Forecasting: from Sentiment-aware News Summarization to Time Series Fusion

牟, 尚決

(Degree)

博士 (工学)

(Date of Degree)

2025-09-25

(Date of Publication)

2027-09-25

(Resource Type)

doctoral thesis

(Report Number)

甲第9331号

(URL)

<https://hdl.handle.net/20.500.14094/0100498371>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



(別紙様式 3)

論文内容の要旨

氏 名 _____ 牟尚決 _____

専 攻 _____ 情報科学 _____

論文題目 (外国語の場合は, その和訳を併記すること。)

Multimodal Deep Learning for Financial Forecasting:
From Sentiment-aware News Summarization to Time
Series Fusion (金融予測におけるマルチモーダル深層学
習: 感情認識型ニュース要約から時系列データの統合まで)

指導教員 _____ 滝口 哲也 _____

(注) 2, 000 字~4, 000 字でまとめること。

This dissertation investigates natural language processing (NLP) and multimodal deep learning techniques for financial analysis, with a focus on methods and technical innovations. Key financial tasks addressed include automatic summarization of news, sentiment analysis of market news, and time series forecasting of economic indicators, all integrated to enhance the understanding and prediction of market trends. The work emphasizes novel approaches to fuse textual and numerical data, aiming to improve the accuracy and interpretability of financial predictive models.

In recent years, NLP has come to play an increasingly pivotal role in the financial domain, as institutions grapple with a deluge of unstructured text—from news articles and earnings reports to social media posts—capable of swaying market sentiment and investment decisions. This accelerating adoption of NLP in finance is fueled by the exponential growth of digital financial communications, heightened regulatory pressures for comprehensive information analysis and oversight, and rapid advancements in large language models (LLMs) that have vastly expanded the scope and accuracy of automated language understanding.

However, applying NLP techniques to financial data presents unique technical challenges, including precise numerical reasoning, integration of heterogeneous and multimodal data sources, and context-aware inference from information that is often fragmented, implicitly conveyed, or embedded with specialized financial terminology and domain-specific knowledge that large language models typically find difficult to accurately capture and interpret. To address these challenges and fully leverage NLP's potential in finance, this dissertation proposes two complementary approaches. First, a sentiment-guided prompt tuning technique is developed to enhance financial text summarization, ensuring that automated summaries capture not only key information but also the underlying sentiment that can influence market reactions. Second, a multimodal learning framework for market forecasting is introduced, combining financial textual data with corresponding time-series data and thereby linking narrative content with quantitative market trends to improve predictive performance.

Chapter 1 introduces the study, highlights its novelty, and presents an outline of this thesis, thereby framing our contributions. Chapter 2 provides a brief overview of NLP applications in the financial domain, establishing the context and significance of the problems addressed. Chapter 3 shows the existing financial summarization and forecasting methods. In this chapter, we discuss the detailed differences between the proposed approach and previous studies and summarize the existing gaps in the current literature to determine our novel contributions. In Chapter 4, we address the task of summarizing financial news articles, emphasizing the importance of accurately capturing sentiment

embedded within news content. We introduce a targeted method for financial news summarization that effectively bridges factual content with sentiment analysis. This approach ensures that automated summaries faithfully represent both the key informational points and the nuanced tone of the original texts, laying essential groundwork for downstream financial applications, where sentiment understanding is as critical as factual accuracy. Having established a robust method for sentiment-aware summarization and interpretation of financial news, we next explore forecasting future financial trends by integrating summarized textual data with numerical time series.

In Chapter 5, we focus on the challenge of accurately forecasting financial time series data—such as stock prices, volatility indices, and economic indicators—by effectively integrating multimodal inputs, specifically numerical time-series data and relevant textual information. To address this, we propose a novel multimodal interpretable Transformer architecture named MM-iTransformer, which is built upon the Transformer framework, recognized for its capability in sequence modeling and attention-based information fusion. The MM-iTransformer model employs two parallel input streams: one capturing structured numerical pattern from time-series data, and the other encoding unstructured textual data. By combining quantitative and qualitative information within a unified model, MM-iTransformer significantly enhances forecasting accuracy and interpretability, providing clear insights into how textual narratives influence numerical market dynamics.

In Chapter 6, we further validate the effectiveness of our proposed forecasting approach, which integrates textual and time-series data, by introducing a multimodal framework that combines numerical financial time-series data with external textual information. Specifically, we present FinTime-Fuse, a forecasting model that employs a fusion mechanism based on LLMs to integrate textual information with quantitative market data. Unlike conventional approaches that rely solely on textual information or on simple similarity-based retrieval of numerical data, our method harnesses the capacity of LLMs to capture intricate correlations between textual information and price fluctuations. The main contribution of this chapter is to demonstrate that LLMs—traditionally employed for conversational tasks—can be successfully applied to multimodal financial forecasting. Furthermore, the proposed approach improves predictive performance, particularly when textual data are limited, and provides clearer insight into how textual information influences financial time series. Notably, our method does not require fine-tuning the LLM; instead, it leverages the model's pretrained knowledge to enable efficient fusion of multimodal representations.

Finally, Chapter 7 draws the conclusions for this thesis.

氏名	牟 尚決		
論文 題目	Multimodal Deep Learning for Financial Forecasting: from Sentiment-aware News Summarization to Time Series Fusion (金融予測におけるマルチモーダル深層学習：感情認識型ニュース要約から時系列データの統合まで)		
審査委員	区 分	職 名	氏 名
	主 査	教授	滝口 哲也
	副 査	教授	羅 志偉
	副 査	教授	太田 能
	副 査		
	副 査		
要 旨			
本研究では、金融分野における自然言語処理とマルチモーダル深層学習の応用、とりわけ「金融テキストの理解」と「マルチモーダル時系列予測」という2つの主要なタスクに焦点を当てている。これらのタスクをとおして、「予測性能を高めるために、テキスト情報をいかに精密に扱い、調整し、活用するか」という、金融シナリオにおける重要な課題の一つに体系的に答えようとしている。 近年の大規模言語モデル（LLM）の急速な発展により、膨大で非構造的な金融テキスト（例えば、ニュースレポート、政策声明、世論など）を扱うために、LLM の導入が金融領域でも進んでいる。しかし、金融特有の高度に専門化された言語や、データのマルチモーダルな異質性、複雑な推論要求により、従来の自然言語処理では、高精度な意味理解や予測推論を実現することは困難である。 金融市場の変動は、多くの場合、ニュースや発表、政策声明などの外部からの情報と、価格や取引量、マクロ経済指標といった内部の数値的な要因が組み合わさって生じる。しかし、従来の統計モデルや深層学習の手法は、数値データの時系列だけに依存しているため、ニュースなどの出来事によって突然起こる変化を十分に捉えることができない。一方で、自然言語処理単独のアプローチでは、価格変動の定量的な制約を考慮できない。 こうした課題を解決するために、本研究では、次の3つのステップからなるアプローチを構築している。まず、文章に含まれる感情や評価の傾向（センチメント）を反映した要約を作成し、次に、文章データと数値データの内容を精度よく対応させる仕組み（バイモーダルアテンションによる整合）を導入する。最後に、大規模言語モデルを使って、それら異なる種類の情報を1つの共通空間にて扱えるようにする仕組みを構築している。これにより、文章の感情的な意味合いの維持、異なる種類の情報の整合性の確保、複雑な入力データの関係性の特徴表現を同時に実現し、より正確で柔軟な予測モデルの構築を可能にしている。 この目的に向けて、本論文では以下の2つの方向から体系的な解決策を提案している。 - 感情整合性を考慮した金融要約手法：金融ニュースの要約生成に感情認識を組み込み、感情キーワードに基づいた軽量のパラメータ調整を行うことで、ニュースの感情的傾向を保ちつつ、正確性も確保し、投資判断への実用性を高める。 - テキストと時系列予測のマルチモーダル融合フレームワーク：Transformer および LLM を活用した2つの融合アーキテクチャを設計し、金融テキストと過去の価格データを効果的に統合する。これにより、金・為替・ビットコインなど複数の金融市場データセットにおいて、予測精度と解釈性の向上を実現する。 本論文の構成は以下のとおりである。 第1章では、研究の背景と動機を説明し、論文全体の構成を述べている。 第2章では、金融における自然言語処理の応用とその課題をレビューし、テキスト感情分析およびマルチモーダルモデリングの技術的な発展についてまとめている。 第3章では、これまでに提案されてきた金融分野の要約作成手法や、時系列データに基づく予測手法を整理している。そのうえで、センチメント、すなわち文章に込められた感情や評価を正しく捉えることの難しさや、文章データと数値データを高精度に対応させるモーダル整合に関する課題を明らかにし、これらの問題を解決するための改善の方向性を示している。			

氏名	牟 尚決
第4章では、記事（文章）の感情的な傾向（センチメント）を崩さずに要約できる、金融ニュース向けの要約手法 PREFIX-PA を提案している。まず、金融分野に特化した言語モデルである FinBERT を使って、ニュース記事の内容が「Positive」「Negative」「Neutral」のどれに当たるかを判定し、その結果を、言語モデルに入力するため「Negative なトーンで要約してください」などの自然な言葉による指示文に変換する。次に、記事（文章）の感情を表す言葉（センチメントワード）と、重要な情報を示すキーワード（情報キュー）を組み合わせ、要約の手がかりとなる「拡張キュー」を作成する。このキューを入力として、事前に学習された言語モデル BART に学習可能な「接頭辞（prefix）」を加え、要約を生成する。BART 本体は変更せず、接頭辞だけを調整することで効率的に学習を行うことができる。ゴールド価格に関するニュースを使った実験では、さまざまなキューの使い方、接頭辞の長さ、損失の重みなどを変えて精度を比較し、要約の質（ROUGE スコア）と感情の一貫性の両面から提案手法の効果を検証している。 第5章では、MM-iTransformer と呼ばれるマルチモーダル・アテンション・フュージョン手法を提案している。この手法では、ニュース記事のテキスト情報を追加の変数として扱い、価格の時系列データと組み合わせることで、文章と数値データの間の関係を深く捉えるマルチモーダルな注意機構（自己注意）を実現している。これにより、市場の動き（金融ダイナミクス）と、ニュースなどの外部情報との複雑な関係性をモデル化できる。具体的には、ニューステキストは金融特化型モデル FinBERT を使って数値ベクトルに変換し、価格データは、各数値項目の時系列を個別の「トークン」として処理する。そのうえで、テキストと価格といった複数の情報に対してマルチアテンションを適用し、異なる変数間の関連性を学習する。最終的には、これらの情報をもとに、将来の価格の予測出力が得られる。実験の結果、MM-iTransformer により、ニューステキストを使わずに数値データだけで学習する従来の時系列モデルに比べて、予測の精度が大きく向上することを示した。 第6章では、金融テキストと時系列データを高度に統合して予測精度を高める新しいクロスモーダル手法 FinTime-Fuse を提案している。これは、従来の手法では実現が難しかったニュースと価格データのきめ細かな連携を、大規模言語モデル（LLM）の構造を活用することで可能にした点が大きな特徴である。本手法では、過去の価格データをパッチに分割し、ニュースのテキスト内容を「キー」として両者の関係をモデル内部で柔軟に結びつける。さらに、金融用語の意味を正確に反映したテキストの情報を、LLM 空間に統合することで、定量データと定性情報の間にあるギャップを埋める手法を実現している。これにより、テキストと数値を単純に並べるのではなく、深い意味レベルで結合し、将来の価格を高精度に予測することが可能となっている。実験では、FinTime-Fuse が従来の数値データ単独モデルやテキスト単独モデルよりも一貫して高い性能が示されていて、金融予測におけるマルチモーダル統合の新しい方向性を示している。 第7章では、本研究全体をまとめ、提案手法の実用的な価値を評価するとともに、今後の研究の方向性について考察している。 以上のように、本研究は、金融分野における自然言語処理とマルチモーダル深層学習の応用、とりわけ「金融テキストの理解」と「マルチモーダル時系列予測」という主要なタスクに関して、3つの課題：センチメントに基づいたテキスト要約、文章データと数値データの内容を精度よく対応させる仕組み（バイモーダルアテンションによる整合）、大規模言語モデルを用いた異なる種類の情報を1つの共通空間にて扱えるようにする仕組みについて研究したものであり、金融予測におけるマルチモーダル深層学習アプローチについて重要な知見を得たものとして価値ある集積である。提出された論文はシステム情報学研究科学位論文評価基準を満たしており、学位申請者の牟 尚決は、博士（工学）の学位を得る資格があると認める。	