LLMFactor: Extracting Profitable Factors through Prompts for Explainable Stock Movement Prediction

Meiyun Wang *  
The University of Tokyo  
omiun20@g.ecc.u-tokyo.ac.jp

Kiyoshi Izumi  
The University of Tokyo  
izumi@sys.t.u-tokyo.ac.jp

Hiroki Sakaji  
Hokkaido University  
sakaji@ist.hokudai.ac.jp

Abstract

Recently, Large Language Models (LLMs) have attracted significant attention for their exceptional performance across a broad range of tasks, particularly in text analysis. However, the finance sector presents a distinct challenge due to its dependence on time-series data for complex forecasting tasks. In this study, we introduce a novel framework called LLMFactor, which employs Sequential Knowledge-Guided Prompting (SKGP) to identify factors that influence stock movements using LLMs. Unlike previous methods that relied on keyphrases or sentiment analysis, this approach focuses on extracting factors more directly related to stock market dynamics, providing clear explanations for complex temporal changes. Our framework directs the LLMs to create background knowledge through a fill-in-the-blank strategy and then discerns potential factors affecting stock prices from related news. Guided by background knowledge and identified factors, we leverage historical stock prices in textual format to predict stock movement. An extensive evaluation of the LLMFactor framework across four benchmark datasets from both the U.S. and Chinese stock markets demonstrates its superiority over existing state-of-the-art methods and its effectiveness in financial time-series forecasting.

1 Introduction

Artificial Intelligence (AI) has become a staple in the financial sector, addressing various challenges such as predicting stock movements (Patel et al., 2015; Henrique et al., 2019), providing robo-advisory services (Xue et al., 2018; Bertrand et al., 2023), and managing risks (Abballi et al., 2022; Wang et al., 2021). Among these tasks, forecasting stock trends is particularly significant, as it leverages historical data to shape trading strategies and pinpoint opportunities for buying or selling stocks.

*Corresponding Author

The Efficient Market Hypothesis (EMH), proposed by Eugene Fama (Fama, 1970), suggests that stock prices reflect all available information, making it difficult to predict future price movements. However, subsequent research has identified limitations to market efficiency, highlighting how phenomena such as information asymmetry (Timmermann and Granger, 2004) and irrational behaviours (Lo, 2004) can induce departures from perfect efficiency. These observations have paved the way for researchers to seek excess market returns by identifying and leveraging market inefficiencies.

Building on this foundation, there is a growing interest in exploring various data types to enhance prediction capabilities. Several studies underscore the importance of news related to stocks in uncovering fundamental market insights (Vargas et al., 2018; Li et al., 2021), while others highlight the significance of understanding the interconnections between companies and industries (Feng et al., 2019; Hsu et al., 2021). Recent research has empirically demonstrated the impact of public sentiment on market trends, with efforts to extract sentiment and keyphrases from news and social media data (Nguyen et al., 2015; Hao et al., 2021).

However, as illustrated in Figure 1, these meth-
ods encounter various limitations. We propose a novel task to enhance stock movement prediction through the use of “factors.” Figure 1 delineates the distinctions among keyphrases, sentiment words, and factors generated by LLMs. For the given stock-related news, keyphrases may outline the content but do not directly correlate with the stock price. Conversely, sentiments are associated with stock price but lack clarity for human interpretation. Consequently, factors offer superiority in three aspects: news explainability, human readability, and stock price explainability.

To incorporate factors into financial forecasting, we introduce a new framework LLMFactor, which derives factors from LLMs through Sequential Knowledge-Guided Prompting (SKGP) and then explains stock price trends. LLMFactor represents a holistic approach to integrating LLMs into financial applications. Initially, our proposed SKGP strategy prompts LLMs to generate factors related to stocks. SKGP starts with a fill-in-the-blank technique to elicit background knowledge (e.g., company relations) related to the stock. Subsequently, we prompt LLMs to identify reliable factors from news articles. Finally, we compile the text-formatted time-series data with these factors and knowledge to predict stock price trends. Experiments across four datasets indicate that LLMFactor yields superior predictive results and can clarify the principles behind these predictions. Additionally, the obtained factors significantly help to dynamically present market changes over time in a format that is accessible and understandable to humans.

The main contributions of this study are summarized as follows:

- **New Task - Factor Extraction**: We introduce a novel task, factor extraction, aimed at extracting significant factors from textual data to assist in predicting time-series data. This task surpasses traditional keyphrase-based and sentiment-based methods in three key areas: the explainability of news content, the readability for humans, and the ability to explain stock price movements.

- **New Strategy - SKGP**: We propose Sequential Knowledge-Guided Prompting (SKGP) as an innovative strategy for leveraging background knowledge in the process of stock movement prediction. Unlike basic prompting methods, SKGP employs a fill-in-the-blank approach, utilizing minimal background knowledge to enhance the richness of the prompt template.

- **New Framework - LLMFactor**: The LLMFactor framework, developed in this study, utilizes factors derived from LLMs to elucidate the dynamic temporal changes. This framework has proven to be a valuable tool in financial applications, providing deep insights into market trends and facilitating factor analysis.

2 Related Work

2.1 Stock Movement Prediction using Textual Data

With the advancement of natural language processing (NLP) techniques, many researchers leverage textual data to forecast stock market trends. (Xu and Cohen, 2018) uses tweets and historical prices to make temporally dependent predictions from stock data. (Luo et al., 2023) models the multimodality between financial text data and causality-enhanced stock correlations. These efforts aim to develop sophisticated models to enhance the prediction accuracy and consider the textual data as a whole.

Furthermore, certain studies focus on extracting more granular insights from textual data. (Zhou et al., 2021) identifies corporate events as the driving force behind stock movements while (Wan et al., 2021) demonstrates the significant association between strong media sentiment and market return. (Wang et al., 2020) enhances stock movement prediction by including expert opinions aggregated from various sources. These insights offer greater clarity to the understanding of market dynamics.

2.2 Time-series Forecasting with LLMs

LLMs, such as variants of the Generative Pre-trained Transformer (GPT), are constructed from wide-ranging and diverse datasets, significantly enriching their extensive knowledge base (Brown et al., 2020). Yet, their foundational structure, grounded in the transformer architecture, is not intrinsically designed for analyzing time-series data (Gruver et al., 2023). This limitation has sparked increased interest among scholars in exploring methods to tailor LLMs for proficient time-series prediction.

(Xue and Salim, 2023) proposes a prompt-based approach to transform numerical inputs and outputs into textual prompts, framing the forecasting task as a sentence-to-sentence conversion. This method allows language models to be directly applied to
Figure 2: LLMFactor: From Sequential Knowledge-Guided Prompting (SKGP) to factor analysis in the stock market. The SKGP technique comprises three steps that sequentially generate knowledge and predict stock movement, guided by this knowledge. In applying factor analysis to stock movements, factors offer timely and concise explanations for their changes.

forecasting tasks. For financial time series forecasting, (Yu et al., 2023) employs prompts to generate summaries and keyphrases from LLMs, incorporating multiple data sources to enhance time-series forecasting. Their approaches offer valuable insights into stock market trends. However, their prompts contain excessive information, resulting in responses from LLMs that are relatively lacking in detail.

2.3 Prompt Engineering

Prompt engineering allows LLMs to address a variety of tasks effectively through carefully crafted prompts. However, identifying the optimal prompt for each specific task is a challenge (Liu et al., 2023). This challenge leads to the exploration of diverse prompting strategies. For instance, (Wei et al., 2022) introduces a chain of thought (CoT) approach, which involves a sequence of intermediate reasoning steps to enhance the capability of LLMs in complex reasoning tasks. Meanwhile, (Liu et al., 2022) proposes a generated knowledge prompting technique, which involves generating knowledge through a language model and then using this generated knowledge as supplementary input for answering questions. Another popular technique, known as Retrieval-Augmented Generation (RAG) (Lewis et al., 2020), incorporates external knowledge into LLMs through retrieve-augment-generate flows. Inspired by these techniques, we propose a novel Sequential Knowledge-Guided Prompting, designed to enhance the accuracy of financial forecasts.

3 LLMFactor

The proposed LLMFactor is a comprehensive framework to predict and explain stock market trends as shown in Figure 2.

3.1 Task Definition

For a given stock, denoted as \( \text{stock}_{\text{target}} \), we consider its associated news published on the target prediction date \( \text{date}_{\text{target}} \), referred to as \( \text{news}_{\text{target}} \), and its historical stock price sequence \( P = \{P_1, P_2, ..., P_t\} \), where \( t \) represents the window size. The task of predicting stock movement is formulated as a binary classification problem, where the stock price sequence is transformed into a series of stock movements, \( \hat{P} = \{\hat{P}_1, \hat{P}_2, ..., \hat{P}_t\} \). In this series, \( \hat{P}_i = 1 \) indicates a rise in the stock price from \( P_{i-1} \) to \( P_i \), whereas \( \hat{P}_i = 0 \) indicates a fall. Our objective is to predict \( \hat{P}_{t+1} \) given \( \text{date}_{\text{target}}, \text{news}_{\text{target}}, \) and \( P \).

3.2 Sequential Knowledge-Guided Prompting

3.2.1 Matching and Acquiring News Background Knowledge

The foundation of our approach is the Sequential Knowledge-Guided Prompting (SKGP) strategy, which comprises three primary stages. Consider Nvidia (NVDA) as our example \( \text{stock}_{\text{target}} \), the initial phase involves matching the stock with relevant news and acquiring background knowledge.
Let \( S = \{(C_i, T_i, I_i) \mid i \in \mathbb{N}, i \leq n\} \) be the stock list, where each tuple \((C_i, T_i, I_i)\) consists of a company \(C_i\), its ticker symbol \(T_i\), and its industry \(I_i\). These tuples are collected from the NYSE and NASDAQ exchanges, with \(n\) being a positive integer representing the total number of such tuples. We match \(S\) with \(\text{news}_{\text{target}}\) to obtain \(\text{stock}_{\text{match}} = \{S \cap \text{news}_{\text{target}}\}\). Following this, we prompt LLMs to obtain relations between \(\text{stock}_{\text{target}}\) and \(\text{stock}_{\text{match}}\). The prompting method is defined as \(\text{LLM} : \text{String} \to \text{String}\), where the input is a \(\text{RelationTemplate}\) “Please fill in the blank: \(\text{stock}_{\text{target}}\) and \(\text{stock}_{\text{match}}\) are most likely in a ___ relationship,” and the output is the type of relationship, denoted as \(\text{relation}\).

This method of deriving background knowledge about \(\text{news}_{\text{target}}\) significantly improves our understanding of the news content. For example, if \(\text{stock}_{\text{target}}\) and \(\text{stock}_{\text{match}}\) are identified as competitors, information about \(\text{stock}_{\text{match}}\) could potentially have a negative impact on \(\text{stock}_{\text{target}}\). Furthermore, numerous studies have underscored the critical role of company relationships in predicting stock market movements (Hsu et al., 2021; Ang and Lim, 2022). Therefore, our fill-in-the-blank technique aims to constrain the response format, facilitating a direct and unambiguous identification of the relationship.

### 3.2.2 Generating Factors that May Affect Stock Prices

The next step for SKGP involves generating factors from \(\text{news}_{\text{target}}\). The importance of these factors is threefold: 1) They are more closely associated with stock movements than keyphrases, sentiments, news summaries, or entire news articles, thus providing a higher likelihood of profitable market trend predictions. 2) Factors derived from news texts offer more immediate and detailed insights into stock price fluctuations compared to those obtained from other sources. 3) They improve the explainability of stock price trends and the rationale behind LLMs’ predictions.

To generate reliable factors, we instruct LLMs to analyze the news content and identify factors that could influence the stock price. This approach fully utilizes the intrinsic knowledge of LLMs. Considering the top \(k\) factors we aim to extract for the given \(\text{stock}_{\text{target}}\) and \(\text{news}_{\text{target}}\), the prompting method is described as \(\text{LLM}(\text{FactorTemplate}) = \text{factor}\), where \(\text{FactorTemplate}\) is a structured sentence: “Please extract the top \(k\) factors that may affect the stock price of \(\text{stock}_{\text{target}}\) from the following news,” followed by \(\text{news}_{\text{target}}\), and the output is the factors generated by the LLMs. The factors produced by LLMs are not restricted to the words found in the news; instead, LLMs consider the content of the news and its potential impact on stock movement, often summarizing the significant elements in the content. For example, the factor “Nvidia stock gain in January” mentioned in Figure 2 suggests that the event was Nvidia’s past stock performance, and the impact was a gain, which could positively influence Nvidia’s future stock price.

#### 3.2.3 Predicting Stock Price Movements

To predict stock movement, we integrate news background knowledge and factors to guide LLMs. Meanwhile, we transform time-series data into a textual format for LLMs to understand. In a sequence of stock movements denoted by \(P = \{P_1, P_2, \ldots, P_t\}\), we introduce a function where \(f(\hat{P}_i = 0)\) is assigned the value “fell” and \(f(\hat{P}_i = 1)\) is assigned “rose,” thereby converting \(\hat{P}\) into a sequence of outcomes \(\text{TextMovement} = \{f(\hat{P}_i) \mid i \in \{1,2,\ldots,t\}\}\), if \(\hat{P}_i = 0\) and “rose” if \(\hat{P}_i = 1\). Given the textual stock movement sequence \(\text{TextMovement}\) and its date series \(\text{date} = \{\text{date}_1, \text{date}_2, \ldots, \text{date}_t\}\), the past stock price movements are transformed into a \(\text{TimeTemplate}\), structured as “On \(\text{date}_i\), the stock price of \(\text{stock}_{\text{target}}\) \(\text{f}(\hat{P}_i)\).”

Subsequently, we construct a \(\text{PriceTemplate}\) that includes an initial instruction, “Based on the following information, please judge the direction of the stock price as rise or fall, fill in the blank and give reasons,” followed by a concluding instruction, “On \(\text{date}_i\), the stock price of \(\text{stock}_{\text{target}}\) will ___.” By integrating the \(\text{relation}, \text{factor}, \text{TimeTemplate}, \) and \(\text{PriceTemplate}\), we articulate the prompting method as \(\text{LLM}(\text{relation}, \text{factor}, \text{TimeTemplate}, \text{PriceTemplate}) = \text{prediction}\). The \(\text{prediction}\) outcome specifies whether the stock price will “rise” or “fall,” along with the rationale for this inference. More details of templates are in Appendix A.

### 3.3 Factor Analysis in the Stock Market

The SKGP presents a powerful technique for predicting stock movements, and the factors derived from SKGP offer additional insights into stock market trends. For instance, as illustrated in Figure 2,
factor analysis can be applied to the stock market. Taking Nvidia’s stock price trend as an example, after a consistent increase over the past five days, the day highlighted with a blue box also exhibits an upward movement. To explain this occurrence, our LLMFactor identifies a concise set of factors, such as “Nvidia stock gain in January, new product announcements, and selection of Nvidia Drive Thor by EV makers.”

4 Experiments

By conducting experiments on benchmark datasets, we aim to address the following research question: Does LLMFactor offer improved prediction accuracy and more insightful explanations for stock market movements?

4.1 Datasets

We evaluate LLMFactor on four benchmark datasets, in alignment with the current SOTAs: a) StockNet (Xu and Cohen, 2018) includes 87 stocks from 9 industries, accompanied by stock-related tweets and historical price data from 2014-01-01 to 2016-01-01, in the US stock market; b) CMIN-US (Luo et al., 2023) comprises the top 110 stocks along with their tweets and historical price data from 2018-01-01 to 2021-12-31, in the US stock market; c) CMIN-CN (Luo et al., 2023) consists of the 300 stocks in the CSI300 index, with their tweets and historical price data from 2018-01-01 to 2021-12-31, in the Chinese stock market; d) EDT (Zhou et al., 2021) includes 54,080 news articles from 2020-03-01 to 2021-05-06, along with related stock and stock price information in the US stock market. Unlike StockNet, CMIN-US, and CMIN-CN, which focus on time-series forecasting, the EDT dataset is centred on news content. The statistics of all datasets are summarized in Table 1.

### Table 1: Statistics of four benchmark datasets

<table>
<thead>
<tr>
<th>Statistics</th>
<th>StockNet</th>
<th>CMIN-US</th>
<th>CMIN-CN</th>
<th>EDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>time series &amp; text</td>
<td>time series &amp; text</td>
<td>time series &amp; text</td>
<td>text</td>
</tr>
<tr>
<td>Data Resource</td>
<td>price sequence &amp; tweets</td>
<td>price sequence &amp; tweets</td>
<td>price sequence &amp; tweets</td>
<td>prices &amp; news articles</td>
</tr>
<tr>
<td>Data Size</td>
<td>19,318</td>
<td>83,553</td>
<td>198,781</td>
<td>54,080</td>
</tr>
<tr>
<td>Stock Market</td>
<td>US</td>
<td>US</td>
<td>CN</td>
<td>US</td>
</tr>
<tr>
<td>Stock Number</td>
<td>87</td>
<td>110</td>
<td>300</td>
<td>4,228</td>
</tr>
<tr>
<td>Date Range</td>
<td>2014-01-01 to 2016-01-01</td>
<td>2018-01-01 to 2021-12-31</td>
<td>2018-01-01 to 2021-12-31</td>
<td>2020-03-01 to 2021-05-06</td>
</tr>
</tbody>
</table>

4.2 Evaluation Metrics

Following previous studies by (Xu and Cohen, 2018; Luo et al., 2023; Zhou et al., 2021), we apply Accuracy (ACC) and Matthews Correlation Coefficient (MCC) as the evaluation metrics. Given the confusion matrix \[
\begin{bmatrix}
tp & fp \\
fn & tn
\end{bmatrix}
\]
ACC and MCC are defined as:

\[
ACC = \frac{tp + tn}{tp + tn + fp + fn}
\]

\[
MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}
\]

4.3 Baselines

**Keyphrase-based.** We employ several models to identify keyphrases in text for prediction purposes:

- **PromptRank (Kong et al., 2023):** An unsupervised approach that utilizes pre-trained language models (PLMs) with a focus on leveraging prompt-based techniques for keyphrase extraction.
- **KeyBERT (Grootendorst, 2020):** Leverages BERT embeddings to identify key sub-phrases in documents through cosine similarity, emphasizing the most representative phrases.
- **YAKE (Campos et al., 2020):** A lightweight unsupervised method for keyword extraction, employing statistical features from individual documents to pinpoint the most pertinent keywords.
- **TextRank (Mihalcea and Tarau, 2004):** An unsupervised method for keyword extraction, employing statistical features from time-series forecasting.
- **TopicRank (Bougouin et al., 2013):** Groups similar expressions into topics and ranks these topics to identify keyphrases.
- **SingleRank (Wan and Xiao, 2008):** An extension of TextRank that weights edges in the co-occurrence graph based on the number of co-occurrences.
• **TFIDF** (Sparck Jones, 1972): A traditional unsupervised approach that evaluates the importance of a word within a document in a corpus.

**Sentiment-based.** We employ various models to analyze sentiment in texts and predict stock market movements:

• **EDT** (Zhou et al., 2021): Utilizes a bi-level event detection framework based on BERT to classify corporate events impacting stock prices.

• **FinGPT** (Zhang et al., 2023): Employs instruction tuning on LLMs to enhance financial sentiment analysis capabilities.

• **GPT-4-turbo**: Builds upon GPT-4’s capabilities, enhancing efficiency and performance, especially for processing longer texts.

• **GPT-4**: Represents a significant leap forward from previous iterations, with improved processing speeds and enhanced handling of extended text inputs.

• **GPT-3.5-turbo**: Known for its comprehensive understanding and human-like text generation, with optimized processing for multiple tasks.

• **RoBERTa**: Applies Financial RoBERTa\(^2\) for sentiment analysis in English texts and a fine-tuned sentiment-specific RoBERTa model\(^3\) for Chinese texts.

• **FinBERT**: A BERT model fine-tuned for financial sentiment classification, available in both English (Araci, 2019) and Chinese versions\(^3\).

**Time-based.** We employ models that leverage both textual and time-series data to predict stock movements:

• **CMIN** (Luo et al., 2023): An end-to-end deep neural network that models the multimodality between financial text data and causality-enhanced stock correlations.

• **StockNet** (Xu and Cohen, 2018): A deep generative model that jointly exploits textual and price signals for stock prediction.

### 4.4 Implementation Details

For keyphrase-based methods, the datasets are divided into two subsets based on labels: one containing data labeled with 0 (indicating a fall) and the other with data labeled with 1 (indicating a rise). Keyphrase extraction is applied to each subset to identify negative and positive keyphrases for each stock. The score for a given text is calculated as follows: Let \(POS\) be the set of positive keyphrases and \(NEG\) be the set of negative keyphrases. For a piece of text \(T\), the score \(S\) is computed as:

\[
S = \sigma \left( \sum_{p \in POS} I(p \in T) - \sum_{n \in NEG} I(n \in T) \right)
\]

where \(\sigma(x)\) is the sigmoid function defined as \(\sigma(x) = \frac{1}{1 + e^{-x}}\), and \(I(\cdot)\) is an indicator function that equals 1 if the condition is true and 0 otherwise.

\[^2\]https://huggingface.co/soleimanian/financial-roberta-large-sentiment
\[^3\]https://huggingface.co/bardsai/finance-sentiment-zh-base

### 4.5 Results

Table 2 presents the results for LLMFactor and various baselines across four datasets. Among all evaluated approaches, our factor-based methods surpass the others, with time-based, sentiment-based, and keyphrase-based methods following in performance. Notably, LLMFactor achieves superior performance over the SOTAs, with improvements of 2.9%, 0.4%, 11%, and 4.8% in MCC across the four datasets, respectively.

The minimal performance differences among keyphrase-based models suggest their limited effectiveness in forecasting stock movements using keyphrases, as keyphrases primarily summarize the main content of the text rather than provide actionable insights on stock prices. Addition-
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyphrase-based</td>
<td>Promptrank (ACL2023)</td>
<td>51.24</td>
<td>0.010</td>
<td>53.28</td>
<td>0.001</td>
<td>50.21</td>
<td>-0.003</td>
<td>51.78</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>KeyBERT</td>
<td>51.95</td>
<td>0.012</td>
<td>53.40</td>
<td>0.009</td>
<td>50.23</td>
<td>-0.004</td>
<td>51.84</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>YAKE</td>
<td>51.91</td>
<td>0.005</td>
<td>53.13</td>
<td>0.001</td>
<td>50.20</td>
<td>0.001</td>
<td>51.88</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>TextRank</td>
<td>51.00</td>
<td>0.003</td>
<td>53.99</td>
<td>0.060</td>
<td>50.38</td>
<td>0.006</td>
<td>51.76</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>TopicRank</td>
<td>51.92</td>
<td>0.008</td>
<td>53.75</td>
<td>0.034</td>
<td>50.26</td>
<td>-0.002</td>
<td>51.80</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>SingleRank</td>
<td>50.32</td>
<td>0.005</td>
<td>53.33</td>
<td>0.004</td>
<td>50.29</td>
<td>0.002</td>
<td>51.85</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>TF-IDF</td>
<td>51.86</td>
<td>0.001</td>
<td>53.71</td>
<td>0.018</td>
<td>50.27</td>
<td>-0.002</td>
<td>51.86</td>
<td>0.017</td>
</tr>
<tr>
<td>Sentiment-based</td>
<td>EDT (ACL2021)</td>
<td>40.31</td>
<td>-0.066</td>
<td>49.86</td>
<td>-0.004</td>
<td>40.00</td>
<td>0.021</td>
<td>75.67</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>FinGPT</td>
<td>54.91</td>
<td>0.083</td>
<td>59.98</td>
<td>0.182</td>
<td>55.78</td>
<td>0.120</td>
<td>53.86</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>GPT-4-turbo</td>
<td>53.56</td>
<td>0.060</td>
<td>64.61</td>
<td>0.284</td>
<td>56.94</td>
<td>0.135</td>
<td>53.57</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>GPT-3.5-turbo</td>
<td>53.88</td>
<td>0.062</td>
<td>62.18</td>
<td>0.260</td>
<td>56.96</td>
<td>0.136</td>
<td>50.94</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>RoBERTa</td>
<td>54.46</td>
<td>0.088</td>
<td>57.75</td>
<td>0.138</td>
<td>52.24</td>
<td>0.064</td>
<td>53.66</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>FinBERT</td>
<td>55.42</td>
<td>0.111</td>
<td>58.26</td>
<td>0.158</td>
<td>55.98</td>
<td>0.121</td>
<td>54.98</td>
<td>0.043</td>
</tr>
<tr>
<td>Time-based</td>
<td>CMIN (ACL 2023)</td>
<td>62.69</td>
<td>0.209</td>
<td>53.43</td>
<td>0.046</td>
<td>55.28</td>
<td>0.111</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>StockNet (ACL 2018)</td>
<td>58.23</td>
<td>0.081</td>
<td>52.46</td>
<td>0.022</td>
<td>54.53</td>
<td>0.045</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Factor-based (ours)</td>
<td>LLMFactor GPT−4-turbo</td>
<td>65.81</td>
<td>0.228</td>
<td>61.71</td>
<td>0.228</td>
<td>60.59</td>
<td>0.245</td>
<td>59.09</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>LLMFactor GPT−3.5-turbo</td>
<td>66.32</td>
<td>0.238</td>
<td>65.26</td>
<td>0.284</td>
<td>57.16</td>
<td>0.196</td>
<td>60.83</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>LLMFactor GPT−4</td>
<td>57.59</td>
<td>0.145</td>
<td>66.42</td>
<td>0.288</td>
<td>56.11</td>
<td>0.139</td>
<td>58.11</td>
<td>0.097</td>
</tr>
</tbody>
</table>

Table 2: Results for LLMFactor and other baselines. The ACC. is shown in percentages (%). Bold text indicates the best results, while underlined text indicates the second-best. We omit the comparison of time-based methods for the EDT dataset due to its lack of time-series data.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>ACC.</th>
<th>MCC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>StockNet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMIN-US</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMIN-CN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: The ablation study results for LLMFactor, summarizing average values across LLMFactor GPT−4-turbo, LLMFactor GPT−3.5-turbo, and LLMFactor GPT−4. The ACC. is expressed in percentages (%), with the highest performance bolded for emphasis. The analysis excludes price data comparisons in the EDT dataset due to the absence of time-series data.

Figure 3: Analyzing the impact of LLMFactor on US and CN markets: LLMFactor is capable of identifying key factors that influence daily fluctuations in stock prices for individual stocks.
ally, sentiment-based methods demonstrate varying performance across different models. LLMs like GPT-4 and GPT-4-turbo excel in identifying sentiment within the text. Although the EDT model achieves the highest accuracy on its specific dataset, its low MCC indicates imbalanced model performance. Unlike sentiment-based methods geared towards efficiently extracting relevant sentiment, time-based methods focus on integrating multiple data types. Since time-based methods consider the text in its entirety without detailed analysis, their performance is on par with that of sentiment-based methods, which may lack temporal context but include detailed textual information. However, our LLMFactor model stands out among other SOTAs due to the SKGP technique, which not only identifies significant factors influencing stock prices from textual data but also incorporates relation and temporal information, thereby enhancing its ability to filter out irrelevant content and provide a more comprehensive analysis.

Analyzing average scores from diverse datasets with the use of three factor-based models, our LLMFactor demonstrates an average accuracy of over 63% and an average MCC of over 0.2 for both the StockNet and CMIN-US datasets, tailored to the US market. However, its performance slightly declines on the CMIN-CN dataset, focused on the CN market, with an average accuracy of 58% and an MCC of 0.19. This variation indicates that the GPT series models are potentially more proficient in processing English-language text. The EDT dataset, which consists solely of stock-related news without historical price data, poses additional challenges, resulting in a marginally lower average accuracy of 59% and an MCC of 0.1 for the LLMFactor. The absence of historical price information in the EDT dataset likely diminishes the LLMFactor’s effectiveness, underscoring the importance of comprehensive data for financial market analysis.

4.6 Data Analysis

4.6.1 Ablation Study

We conduct an ablation study on the LLMFactor to demonstrate the improvement contributed by each layer. The first layer utilizes only the TimeTemplate, which incorporates past price movements. The second layer incorporates additional factors extracted, and the third layer further includes relations obtained. We average the scores for ACC and MCC from $LLM_{GPT-4-turbo}$, $LLM_{GPT-4}$, and $LLM_{GPT-3.5-turbo}$, as shown in Table 3. The price layer accounts for approximately 86% and 32% of the total performance in ACC and MCC, respectively, while the factor layer contributes an improvement of 9% in ACC and 46% in MCC. The relation layer leads to a 5% and 22% enhancement in ACC and MCC, respectively. These results imply that the factor layer contributes most significantly to the overall performance of the proposed LLMFactor.

We also conduct experiments on different types of FactorTemplate, as shown in Appendix C.

4.6.2 Case Study

In this section, we highlight the practical effectiveness of LLMFactor through an illustrative application. Figure 3 presents a case study where we apply factor analysis to selected stocks in the US and CN markets. For instance, on September 17, 2019, Apple Inc.’s stock (AAPL) experienced an increase, attributed to significant developments such as a $250 million investment in its supplier, Corning Incorporated, and the positive initial demand for the iPhone 11. Conversely, on May 8, 2018, Tesla Inc.’s stock (TSLA) experienced a decline, mainly due to a record quarterly loss, growing negative concerns among investors, and the increasing competition in the autonomous and electric vehicle markets. This case study demonstrates LLMFactor’s ability to effectively integrate background knowledge of company affiliations with both historical news and price data. By analyzing these diverse data sources, LLMFactor significantly enhances the interpretability of stock market dynamics.

5 Conclusion

In this study, we introduce LLMFactor, an innovative framework designed to predict and elucidate stock market trends accurately. At the core of LLMFactor is the Sequential Knowledge-Guided Prompting (SKGP) strategy, which integrates background knowledge, stock-related factors, and temporal data to forecast stock movements. Through rigorous testing on four benchmark datasets, LLMFactor has proven its superiority over SOTAs that rely on keyphrases, sentiment analysis, and multimodal data inputs. The application of factor analysis highlights LLMFactor’s novelty and effectiveness, establishing it as a powerful tool for financial analysis. This research represents a significant advancement in leveraging LLMs for transparent and explainable financial forecasting.
6 Limitations

We recognize three main limitations:

• Although LLMFactor’s foundation relies on extensively discussed factors, converting temporal data into text format remains vital for financial forecasting. Future studies could explore additional methods for this conversion and their integration into LLMs.

• Due to the variable nature of LLM responses, replicating our experimental results precisely is challenging. In future research, we aim to explore further methodologies to improve reproducibility.

• Evaluating LLMFactor against benchmark datasets enables comparison with leading models, but it’s essential to carefully assess the quality of factors extracted from texts of various lengths and types. This research marks the beginning of a novel approach to applying factors in finance. Moving forward, we will concentrate on enhancing the quality of these factors.

Acknowledgements

We sincerely thank the anonymous reviewers for their reviews. This work was supported by JST SPRING, Grant Number JPMJSP2108, and JST-PRESTO, Grant Number JPMJPR2267.

References


Ming Xu. 2023. pke_zh: Python keyphrase extraction toolkit for chinese.


A Templates of SKGP

The fundamental templates employed at each step of SKGP are detailed in Table 5 for English and Table 6 for Chinese. In the final step of SKGP, we compile the responses from previous steps to formulate prompts for LLMs. Nonetheless, due to the variable nature of LLM responses, the exact reproducibility of results cannot be guaranteed. Instead, we demonstrate the overall superior performance of the LLMFactor.

B Details of Pre-trained Models

We employ the pke (Boudin, 2016) and pke_zh (Xu, 2023) toolkits to access keyphrase-based methodologies, ensuring alignment with established baselines. Additionally, configuration details for sentiment-based models are outlined in Table 4, provided by HuggingFace. The models specified at the top are designated for English texts, whereas the ones at the bottom cater to Chinese texts.

<table>
<thead>
<tr>
<th>Model</th>
<th>HuggingFace Key</th>
<th>Model Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinBERT</td>
<td>ProsusAI/finbert</td>
<td>438MB</td>
</tr>
<tr>
<td></td>
<td>bankai/finance-sentiment-zh-base</td>
<td></td>
</tr>
<tr>
<td>RoBERTa</td>
<td>IDEA-CCNL/ErLangshen-Roberta-110M-Sentiment</td>
<td>409MB</td>
</tr>
<tr>
<td></td>
<td>onpattern/bianzheng-zh-base</td>
<td>409MB</td>
</tr>
<tr>
<td>FinGPT</td>
<td>oliverwang15/FinGPT_ChatGLM3_Sentiment_Instruction_LoRA_FT</td>
<td>7.89GB</td>
</tr>
</tbody>
</table>

Table 4: Details of pre-trained models

C Comparative Experimental Results of Factor Templates

The experimental results for various factor templates, utilizing three factor-based models: LLMFactorGPT−4−turbo, LLMFactorGPT−4, and LLMFactorGPT−3.5−turbo, are showcased in Table 7 for English templates and in Table 8 for Chinese templates. The datasets tested include CMIN-US and CMIN-CN. Within the English templates, the last template achieves the highest performance, while among the Chinese templates, the initial template emerges as the most effective. The consistent effectiveness of LLMFactor across different templates underscores its robustness in predicting stock market movements.
Step 1
Please fill in the blank and return a complete sentence: stock\textsubscript{target} and stock\textsubscript{match} are most likely in a ___ relationship.

Step 2
Please extract the top \( k \) factors that may affect the stock price of stock\textsubscript{target} from the following news.

Based on the following information, please judge the direction of the stock price from rise/fall, fill in the blank and give reasons.

These are the main factors that may affect this stock’s price recently: factor.

These are the connections between the companies that have appeared in the news: relation.

On date\textsubscript{i-5}, the stock price of stock\textsubscript{target} \( f(P_{i-5}) \).
On date\textsubscript{i-4}, the stock price of stock\textsubscript{target} \( f(P_{i-4}) \).
On date\textsubscript{i-3}, the stock price of stock\textsubscript{target} \( f(P_{i-3}) \).
On date\textsubscript{i-2}, the stock price of stock\textsubscript{target} \( f(P_{i-2}) \).
On date\textsubscript{i-1}, the stock price of stock\textsubscript{target} \( f(P_{i-1}) \).
On date\textsubscript{i}, the stock price of stock\textsubscript{target} will ___.

Table 5: SKGP Templates in English

<table>
<thead>
<tr>
<th>Step</th>
<th>CN-Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step1</td>
<td>请填空并返回完整的句子：stock\textsubscript{target} 和 stock\textsubscript{match} 最可能是___关系。</td>
</tr>
<tr>
<td>Step2</td>
<td>请从以下新闻中提取可能影响 stock\textsubscript{target} 股价的前( k )个因素。</td>
</tr>
</tbody>
</table>

根据以下信息，请判断股票价格是上涨还是下跌，填写在空白处并给出理由。

这些是最近可能影响该股票价格的主要因素：factor
这些是新闻中出现过的公司之间的关系：relation
在 date\textsubscript{i-5}，stock\textsubscript{target} 的股价 \( f(P_{i-5}) \)。
在 date\textsubscript{i-4}，stock\textsubscript{target} 的股价 \( f(P_{i-4}) \)。
在 date\textsubscript{i-3}，stock\textsubscript{target} 的股价 \( f(P_{i-3}) \)。
在 date\textsubscript{i-2}，stock\textsubscript{target} 的股价 \( f(P_{i-2}) \)。
在 date\textsubscript{i-1}，stock\textsubscript{target} 的股价 \( f(P_{i-1}) \)。
在 date\textsubscript{i}，stock\textsubscript{target} 的股价将___。

Table 6: SKGP Templates in Chinese

<table>
<thead>
<tr>
<th>Factor Template</th>
<th>LLMFactor\textsubscript{GPT-4−Turbo}</th>
<th>LLMFactor\textsubscript{GPT−4}</th>
<th>LLMFactor\textsubscript{GPT−3.5−Turbo}</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>MCC</td>
<td>ACC</td>
<td>MCC</td>
</tr>
<tr>
<td>Please extract the top ( k ) factors that may affect the stock price of stock\textsubscript{target} from the following news</td>
<td>61.71 0.228</td>
<td>65.26 0.284</td>
<td>66.42 0.288</td>
</tr>
<tr>
<td>Please identify the primary top ( k ) factors influencing stock\textsubscript{target}'s stock price based on the news provided</td>
<td>66.98 0.292</td>
<td>64.83 0.246</td>
<td>65.26 0.298</td>
</tr>
<tr>
<td>Please analyze the provided news and pinpoint the top ( k ) major factors impacting the stock price of stock\textsubscript{target}</td>
<td>65.29 0.295</td>
<td>69.21 0.312</td>
<td>66.56 0.293</td>
</tr>
</tbody>
</table>

Table 7: Experimental Results for Different Factor Templates in English

<table>
<thead>
<tr>
<th>Factor Template</th>
<th>LLMFactor\textsubscript{GPT-4−Turbo}</th>
<th>LLMFactor\textsubscript{GPT−4}</th>
<th>LLMFactor\textsubscript{GPT−3.5−Turbo}</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>MCC</td>
<td>ACC</td>
<td>MCC</td>
</tr>
<tr>
<td>Please extract the top ( k ) factors that may affect the stock price of stock\textsubscript{target} from the following news</td>
<td>60.59 0.245</td>
<td>57.16 0.196</td>
<td>56.11 0.139</td>
</tr>
<tr>
<td>Please identify the primary top ( k ) factors influencing stock\textsubscript{target}'s stock price based on the news provided</td>
<td>57.92 0.147</td>
<td>64.79 0.109</td>
<td>65.13 0.139</td>
</tr>
<tr>
<td>Please analyze the provided news and pinpoint the top ( k ) major factors impacting the stock price of stock\textsubscript{target}</td>
<td>64.29 0.160</td>
<td>59.33 0.053</td>
<td>59.49 0.033</td>
</tr>
</tbody>
</table>

Table 8: Experimental Results for Different Factor Templates in Chinese