Base On Hierarchical Mechanism Modality Input To Forecast Stock Price

36920221153077Jingyuan Feng, 36920221153157Zhouao Zhou
36920221153156Yikun Zhou, 36920221153068Xunfei Cai

Institute of Artificial Intelligence Xiamen University

Abstract

The stock market is a well-recognized complex dynamic system with many influencing factors such as non-stationarity, non-linearity, high noise, long memory, etc. It is difficult to simply explain it through a mathematical model. Therefore, the analysis and prediction of financial data has been a very challenging task for a long time. In this filed, many deep neural networks methods have received a lot of attention, and they achieve more accurate results than traditional linear and nonlinear methods. Most works begin with two aspects,on the hand, is base on traditional technical and fundamental analysis, utilizing pure stock data(open price ,close price,etc), on the anther hand, Text-mining technologies have substantially affected financial industries. As the data in every sector of finance have grown immensely, text mining has emerged as an important field of research in the domain of finance.some works pay attention on the influence of news on stock prices.Recently, there has been a growing interest in utilizing graph-structured data in stock market,however,it has weakness that is neglects to mine textual information and traditional technical and fundamental analysis information. Therefore, we propose base on hierarchical mechanism modality input to forecast stock price, modality inputs including price data, and text data, and extract information in a hierarchical way to retain more valid information, and finally send into the graph network to predict the rise or fall of a single stock through the correlation between stocks. Compared to the previous model, our accuracy has improved.

1.Introduction

The stock market forecast has been a challenging problem to solve. The efficient-market hypothesis presented by Fama (Fama 1995) suggests that, in the efficient information markets, stock prices behave like a random walk and it is impossible to forecast direction and magnitude changes. He proposed three categories of efficiency: weak form, where past price movements can't be used to predict the future ones;semi-strong form, where neither the past price movements and any public information is relevant for predicting the market; strong-form, where none of the information, public or private can be used to forecast the market. Despite Fama's hypothesis, the scientific community has proposed different ways to forecast stock market,conventional research focused on time series and technical analysis of a stock, i.e., using patterns from historical price signals to forecast stock movements,and deep learning makes it more popular.

but it is far from enough to analyze and predict stock price data only relying on deep learning. We also need to introduce text data as the data for analysis Sources, because some text information can also represent investors' investment preferences, which has a huge impact on stock prices. In the financial trading market, investor sentiment can often represent the investment trend of retail investors. For example, news of the death of Jobs caused Apple's stock price to drop significantly. Therefore, when analyzing stocks, text information is also a data source that cannot be ignored.

some works in stock price forecasting relies on information mining like news reports, tweets, etc.to make stock decisions.however, The amount of information and the quality of the information are critical.Trading decisions made with a small amount of information can easily make investors bear huge risks.On the contrary,Exploding information on the Internet together with the advancing development of natural language processing is also make Investors confused . investors cannot unveil market trends and volatility from online content. Unfortunately, the quality, trustworthiness, and comprehensiveness of online content related to stock market vary drastically, and a large portion consists of the low-quality news, comments, or even rumors. To address this challenge,we propose our model—HMGATs, For the input data -Twitter text, multi- Hierarchical to extract information , and send to two layers of GRUs and attention is used to effectively extract important content and avoid interference from invalid information.

In 1989, Malkiel (Malkiel 1989). proposed the efficient market hypothesis (EMH), which pointed out that financial markets are informationally efficient, the hypothesis states that financial markets are informationally efficient, so we can know that stock prices reflect all information Combined results. Previous studies whether fundamental analysis or technical analysis focused primarily on stock-related datasets, The fundamental analysis of a company involves an in-depth analysis of its performance and profitability. For technicians, stock prices are considered as only typical time series data with complex patterns. With appropriate preprocessing and modeling, patterns can be analyzed, from which profitable patterns may be extracted. The information used for technical analysis consists of mainly closing prices, returns, and volumes.

technical analysis works have focused on how to extract meaningful features from raw price data. In the finance industry, features extracted from such data are called technical indicators and include adaptive moving average, relative strength index, etc.some works have utilized indicators to more accurately predict the movement of stock prices.but lack of analysis of the relationship between social media text and stocks limited their potential to capture broader data that influences stock movements. Our model -HMGATs ,There is not only text input, but also raw data input based on fundamentals. Through one layer of GRUs and one layer of temporal attention, the price vector is obtained.

Both price signals and text information show continuous context dependence, and a single sample can be classified as a unified sequence although it cannot reveal enough information. Different text information has different impacts on stock prices(Hagenau, Liebmann, and Neumann 2013). Some previous studies only focus on the fusion of two modal data and do not care about the connection between stocks. Therefore,Our model -HMGATs makes up for this part. We add graph attention network to explore the correlation between different stocks. And based on and the current stock of the highest degree of correlation to predict the stock price up or down.

On the basis of EMH and previous research, we construct a new network that can perform deep learning from price data, text data and the relationship among stocks. We propose our model -HMGATs,capture the relevant signals in different data through layered time signals, and use this as input data to train the graph attention network GAT, and finally perform binary classification prediction of stock prices.

2.Related work

Some scholars used neural network technology to study financial time series in the 1990s and predicted the daily rate of return of IBM stocks(White 1988). However, due to the gradient explosion problem of the traditional BP neural network, the result will converge to a local minimum(Li et al. 2012). With the advent of the big data era and the widespread application of deep learning, many scholars have also tried to apply the newly proposed recurrent neural network (RNN) model and its improved model LSTM model in financial research. In particular, LSTM has become the first choice of many researchers due to its unique potential in time series modeling(Chen, Zhou, and Dai 2015).

Siami and Namin (Siami-Namini and Namin 2018) have comparatively studied the two financial time series analysis methods, ARIMA and LSTM, and the results show that the LSTM model is 85 % more accurate than the ARIMA model.

Yingying Yan (Yan and Yang 2021) has designed an encoder-decoder model of attention mechanism based on LSTM. Time series forecasting mainly faces two major problems. The first is that the target sequence in the time sequence is affected by multiple input feature sequences. The second is the time correlation of the time series. When using traditional models for analysis, the impact of this interaction will be ignored, resulting in insufficient generalization capabilities of the model. To solve the problems in time series prediction, the attention mechanism can be added to the encoder and decoder parts of the encoder-decoder model. The feature attention mechanism is introduced into the encoder to calculate the attention weight of the input feature at the current moment. This weight indicates the importance of the input feature to the current target task.

Although Transformer-based methods have significantly improved state-of-the-art results for long-term series forecasting, they are not only computationally expensive but more importantly, are unable to capture the global view of time series(Liu et al. 2019). To address these problems, Tian Zhou et al (Zhou et al. 2022) propose to combine Transformer with the seasonal-trend decomposition method, in which the decomposition method captures the global profile of time series while Transformers capture more detailed structures. To further enhance the performance of Transformer for long-term prediction, they exploit the fact that most time series tend to have a sparse representation in well-known basis such as Fourier transform, and develop a frequency enhanced Transformer. Besides being more effective, the proposed method, termed as Frequency Enhanced Decomposed Transformer (FEDformer), is more efficient than standard Transformer with a linear complexity to the sequence length.

Stock predictions are influenced not only by price fluctuations over time series, but also by natural language processing from sources such as policy, news, social media data, and earnings calls. Previous stock forecasting models focused only on price fluctuations and direct economic parameter factors such as GDP on time series, classified as technical analysis (TA) method. The new stock forecasting model is classified as a method called fundamental Analysis (FA)](Huang, Capretz, and Ho 2019), which incorporates factors related to natural language processing in the model from sources other than numbers, such as policy, news and social media data, and earnings calls. Compared with previous models which only consider economic figures, the model based on this method has the advantage of taking into account factors such as the government, the market and the sentiment of Stockholders.

However, the existing stock prediction models based on natural language processing still have some shortcomings, which lack the method to correlate stock price and market sentiment through text clues. These models believe that stock price fluctuations and natural language factors are independent, which is not consistent with the actual situation.

3.Proposed Solution

In theory, stock prices are predictable, but there are many factors that affect them, and until now, their effects on stocks have not been clearly defined. This is because stock prediction is highly nonlinear, which requires the prediction model to be able to deal with nonlinear problems(Shah, Isah, and Zulkernine 2019). In addition, stock has the character-

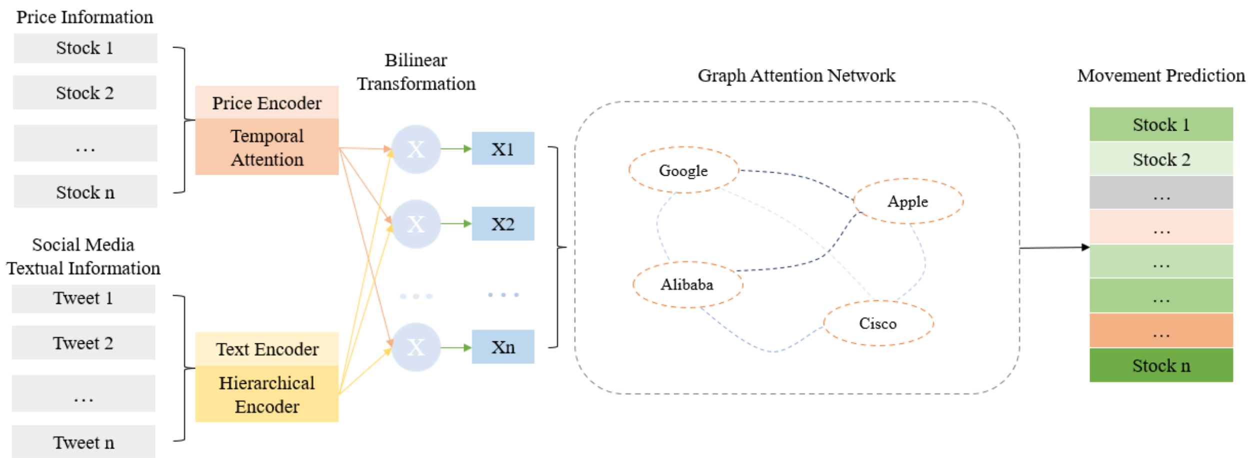

Figure 1: An overview of HMGATs:Encoding Mechanisms, GAT Mechanism, Joint Optimization
.

istics of time series, so it is suitable to use cyclic neural network to forecast stock. Although the cyclic neural network (RNN) allows the persistence of information, the general RNN model is weak in describing the time series data with long memory. When the time series is too long, the RNN training becomes very difficult due to the phenomenon of gradient dissipation and gradient explosion. The Long Short Term Memory (LSTM) model proposed by Hochreiter and Schmidhuber was modified on the basis of RNN structure(Hochreiter and Schmidhuber 1997), which solved the problem that RNN model could not describe the long term memory of time series. In conclusion, the LSTM model in deep learning can describe the long-term memory of time series well.

In natural language processing, public news and social media are the two main information resources. Traditional methods use feature engineering(Kuhn and Johnson 2019), but with the emergence of neural networks, event-driven methods and structured event representation are being studied. More direct use of hierarchical attention mechanism to directly mine the news sequence(Xu et al. 2020). But stock forecasting is still difficult because it is randomly affected by news events, resulting in a random-walk characteristic. Compared with the discriminative model, the generative model has a natural advantage in depicting the generation process from market information to stock signals and in introducing randomness(Leng, Liu, and Guo 2022). However, existing researches on natural language processing have not solved the problem of time dependence between motion prediction(Liu et al. 2021). For example, when a company suffers a major scandal on its first trading day, its share price tends to fall on the following trading day. If the stock forecasting model can predict such a decline pattern, it can better predict all the trends during the emergence of market information. Otherwise, forecast accuracy may be compromised during such periods. Such forecasts, which rely on market information, also show a degree of correlation over time by undoing the effects over time.

In order to solve the above prominent research gaps in predictive modeling related to time series and stochastic market information, this paper proposes a solution to predict stock movements by learning relevant information from Twitter tweets and historical prices, and using corporate relationships between stocks.

We described the stock trend prediction as a binary classification of the rise and fall problem, encoded two kinds of characteristic data of historical stock prices and Twitter tweets, used bilinear transformation to convert these characteristic data into continuous cycle variables for modeling, and let the model directly learn from these data to generate stock movements. At the same time, it can describe the dependence relationship between stocks, provide a multi-task learning objective, and show the correlation between stocks of different companies through the graph attention mechanism. We evaluated the results using a dataset of high-volume stocks collected, and found that the model showed more productive results for forecasting by combining tweets with historical stock price data.

4.Experiments

1.Components and Learning

As shown in Figure 1, the model first encodes the market data of each stock within a certain range, denoted as x_t=B(c_t,q_t), where c_t is the market information extracted within this range, i.e., Twitter messages and other information. q_t is based on historical share price data within this range. Then the c_t and q_t are fused by bilinear transformation to get x_t. Then, the graph attention network is used to map the relationships between the stocks of various companies. Finally, through the joint optimization of GAT and HMGATs based on the propagation characteristics of interpopulation relations, the correlation of market information and historical stock price data to stock prices can be obtained, so as to predict the stock price trend.

2.Price Encoder

As shown in Figure 2, price feature q_t is generated by en-

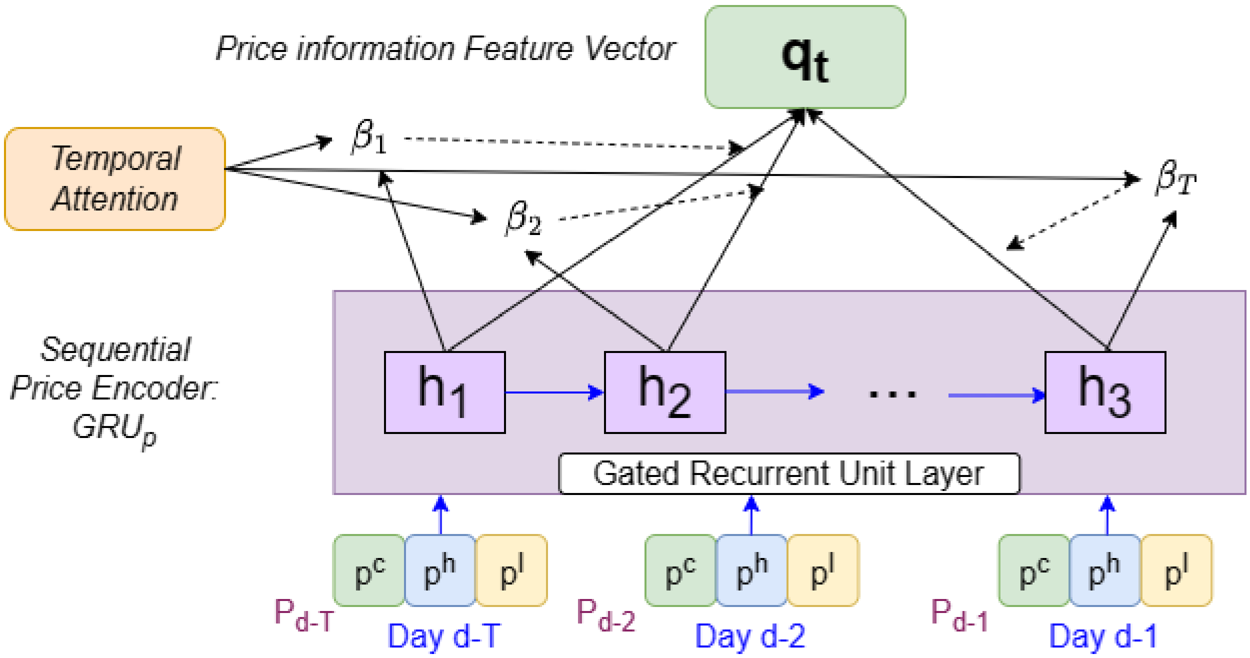

Figure 2: An Overview of the Price Encoder

.

coding historical stock price movements through price encoder. It takes the daily stock price characteristics from the past period of time, encodes the time trend of historical stock prices, and then uses the GRUs to get the sequential dependencies between trading days.

3.Social Media Information Encoder (SMI)

As shown in Figure 3, market information feature ct is generated by encoding relevant tweets of each stock in each trading day through social media information encoder. With the rapid spread of information on the Internet, good or bad tweets about companies sent out by social media platforms such as Twitter can affect shareholders' confidence in holding shares. Tweets are not only a release of corporate information, but also affect the sentiment and direction of stock investment.

4.Blending Multimodal Information After obtaining price feature qt and market information feature c_t, ct and q_t are integrated by bilinear transformation. Signals from different features carry supplementary information about stock price trend factors, and x_t is obtained. Then, the stock and its correlation are modeled graphically using GAT.

5.Dataset and Training Setup

We selected 30 companies with high-volume stocks, and each company selected 508 stock price data for the two-year period from December 24, 2013 to December 31, 2015, for a total of 15,240 stock price data. Since high-volume stocks tend to be discussed more on Twitter, we also got data on eight tweets per day for each company over the two-year period from December 24, 2013 to December 31, 2015, for a total of 121,920 tweets.

6.Results and Analysis

As shown in Table 1, based on these data, we ran 6000 rounds and obtained some data results and their associated effects. The values of F1, accuracy, MCC and LOSS were calculated for 6 times per 1000 rounds.At 1000 rounds, F1 score and Accuracy reached 0.4970, Matthew's Correlation Coefficient (MCC) was -0.0181, and LOSS reached 0.7173. At 2000 rounds, F1 score and Accuracy reached 0.5178, Matthew's Correlation Coefficient (MCC) was 0.0175, and LOSS reached 0.7041. At 3000 rounds, F1 score and Accuracy reached 0.4993, Matthew's Correlation Coefficient (MCC) was -0.0103, and LOSS reached 0.7164. At 4000 rounds, F1 score and Accuracy reached 0.5096, Matthew's Correlation Coefficient (MCC) was -0.0126, and LOSS reached 0.7102. At 5000 rounds, F1 score and Accuracy reached 0.5148, Matthew's Correlation Coefficient (MCC) was -0.0079, and LOSS reached 0.7150. At 6000 rounds, F1 score and Accuracy reached 0.4993, Matthew's Correlation Coefficient (MCC) was -0.0204, and LOSS reached 0.7091. As can be seen from the table, the accuracy is not high and the result is not significant after multiple rounds of execution. However, it can also be shown that it is effective to make stock prediction by integrating stock history price and company Twitter data. The method of combining time series data prediction with natural language processing focuses on considering the connection between social media information and stock fluctuations.

Bilinear transformation is generally better than the traditional series and attention variant. It can better describe the interrelation between different feature signals, so that the model pays more attention to the mutual influence between different feature signals in the learning process. From the

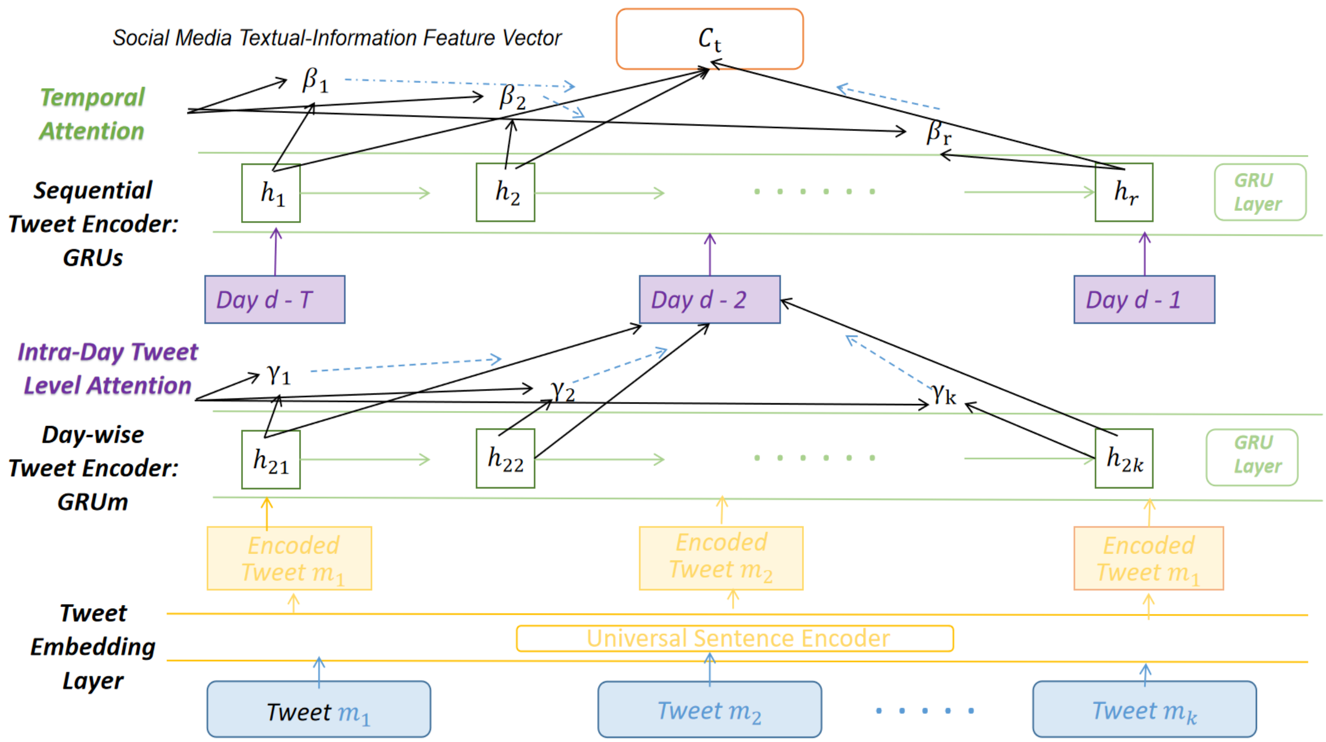

Figure 3: Social Media Information Encoder

.

Epoch	F1	Accuracy	MCC	LOSS
1000	0.4970	0.4970	-0.0181	0.7173
2000	0.5178	0.5178	0.0175	0.7041
3000	0.4993	0.4993	-0.0103	0.7164
4000	0.5096	0.5096	-0.0126	0.7102
5000	0.5148	0.5148	-0.0079	0.7150
6000	0.4993	0.4993	-0.0204	0.7091

Table 1: Results

results, darker tweets or historical prices showed a greater impact on stock price trends. We can see that bilinear transformation integrates the nonlinear relationship between the two signal features, and can capture some more specific features in the results.

7.Qualitative Analysis

We extend the analysis of other application scenarios of the model. It is found that the model can predict the further decline of stock price by relying on the historical stock price and related company tweets, but the two signal features alone cannot accurately predict the rise of stock price. But taking into account the effects of social media data, positive tweets that get more attention after a big drop still indicate that the stock price is likely to rise. In a period of time, there is a strong correlation between the stock prices of some companies, such as companies of the same type. The graph attention mechanism is used in the model to depict the relationship between stocks and show the relationship between the stocks of these companies, indicating that the model may have a wider range of application scenarios.

5.Conclusion

The stock market forecast is related to many factors. In this project, we try to consider more factors in the model. We use multimodal data including historical prices and tweets. Besides that, we extract correlations between stocks to improve the performance of the model. The result shows that using multimodal data can train a better model. In the future, we plan to use news articles, earnings calls, and other data sources to get a better model. Actually, different factors affect the stock market to different degrees, so we should consider different weights. And then we plan to consider more complex corporate relationships. The changes in downstream companies can directly affect upstream companies. Finally, the stock market changes all the time. So we think using reinforcement learning in the model can make it time-sensitive.

References

Chen, K.; Zhou, Y.; and Dai, F. 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE international conference on big data (big data)*, 2823–2824. IEEE.

Fama, E. F. 1995. Random walks in stock market prices. *Financial analysts journal*, 51(1): 75–80.

Hagenau, M.; Liebmann, M.; and Neumann, D. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55(3): 685–697.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Huang, Y.; Capretz, L. F.; and Ho, D. 2019. Neural network models for stock selection based on fundamental analysis. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 1–4. IEEE.

Kuhn, M.; and Johnson, K. 2019. *Feature engineering and selection: A practical approach for predictive models*. CRC Press.

Leng, J.; Liu, W.; and Guo, Q. 2022. Stock movement prediction model based on gated orthogonal recurrent units. *Intelligent Systems with Applications*, 200156.

Li, J.; Cheng, J.-h.; Shi, J.-y.; and Huang, F. 2012. Brief introduction of back propagation (BP) neural network algorithm and its improvement. In *Advances in computer science and information engineering*, 553–558. Springer.

Liu, J.; Lin, H.; Liu, X.; Xu, B.; Ren, Y.; Diao, Y.; and Yang, L. 2019. Transformer-based capsule network for stock movement prediction. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, 66–73.

Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; and Neubig, G. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.

Malkiel, B. G. 1989. Efficient market hypothesis. In *Finance*, 127–134. Springer.

Shah, D.; Isah, H.; and Zulkernine, F. 2019. Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, 7(2): 26.

Siami-Namini, S.; and Namin, A. S. 2018. Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.

White, H. 1988. Economic prediction using neural networks: The case of IBM daily stock returns. In *ICNN*, volume 2, 451–458.

Xu, H.; Chai, L.; Luo, Z.; and Li, S. 2020. Stock movement predictive network via incorporative attention mechanisms based on tweet and historical prices. *Neurocomputing*, 418: 326–339.

Yan, Y.; and Yang, D. 2021. A stock trend forecast algorithm based on deep neural networks. *Scientific Programming*, 2021.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. *arXiv preprint arXiv:2201.12740*.