2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) | 978-1-6654-9260-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IITCEE57236.2023.10091033

Stock Movement Prediction via Temporal Convolutional Network and Interactive Attention Network

1st Sing Guo
School of Information Science and Technology/School of Cyber Security
Guangdong University of Foreign Studies
Guangzhou 510006, China
singguo0919@163.com

2nd Hui Ai
School of Accounting
Guangdong University of Foreign Studies
Guangzhou 510006, China
aihuif@163.com

3rd Shanxin Li
School of Finance
Guangdong University of Foreign Studies
Guangzhou 510006, China
1031369426@qq.com

***Abstract*—Stock movement prediction is an important study that can provide investors with some reliable trading signals and long-term stable investment returns. Although many existing studies show that using text data of social platforms and historical share prices can effectively predict stock movement, the existing methods still have some limitations. For example, (i) future information is inevitably used to obtain the features of text and price; (ii) The interaction between text data of social platforms and historical share price data was not considered. In order to alleviate these restrictions, we propose a novel method to predict the stock movement based on Temporal Convolution Network (TCN) and Interactive Attention Network (IAN), in which TCN can effectively avoid using future information when obtaining the features of text and price. In addition, IAN can get the interactive feature between social texts and historical share prices. In the end, we employ a three-layer neural network classifier to predict the stock movement. The experimental results show that the proposed model can obtain a competitive results compared with other existing methods.**

Keywords—Stock movement prediction, Temporal Convolution Network, Interactive Attention Network

I. Introduction

Stock movement prediction has an important influence and attracted the attention of many researchers in the field of Fintech[1]. However, the stock market is uncertain and highly complex and the stock price movement follows a random walk form[2]. Therefore, it is a challenge to find an advanced prediction method to accurately predict stock movement in a highly random market.

In early research, fundamental analysis and technical analysis are often used to help researchers and investors make decisions in the stock market[3]. Fundamental analysis usually selects some fundamental indicators of companies corresponding to stocks, such as the P/E ratio, P/B ratio, and P/S ratio, etc[3]. Technical analysis selects related to historical stock price fluctuations as analysis indicators, such as OBV, MACD, and RSI, etc[3]. However, they only capture the associations and change rules of historical time series data of stocks, so the prediction accuracy is random and uncertain. In other words, fundamental analysis and technical analysis regard the stock movement prediction as a time series problem, without considering the impact of emergency events and news[4]. Therefore, some researchers employed public news to predict the trend of stocks[5-7]. Unfortunately, most of the public news is lagging behind, and the relevant news is often released after the stock price has an obvious trend. Recently, some researchers have found that social media is more timely than public news, and the study of predicting stock movement via social media texts has attracted more and more researchers' attention[8-10].

With the development of data mining and Natural Language Processing technology(NLP), traditional machine learning and deep learning methods are used to process text information and stock price information to predict stock movement. For example, Kalyani et. al[5] employed traditional machine learning methods, including Random Forest (RF), Support Vector Machine (SVM), and naive Bayes, to analyze the emotion of news texts and predict stock movement via the emotion reflected by news texts. Long et al.[11] used SVM to obtain the structural features between news texts for forecasting stock movement. However, the methods based on traditional machine learning require a lot of feature engineering, which is time-consuming and laborious. Therefore, some researchers began to use deep learning to predict the stock movemet. For instance, Ding et al.[12] used Convolutional Neural Network (CNN) to extract the short-term and long-term impact features of financial news texts to predict stock movement. Hu et al.[1] designed a Hybrid Attention Network (HAN) based on the principles of online news learning, including sequential content dependence, diversity influence, and efficient learning, to learn the temporal features of news texts for stock movement prediction.

With further research, some researchers found that the social media opinion are highly related to the stock price. At the same time, extracting and integrating features from the social media text and the stock price can effectively improve the accuracy of stock movement prediction. For example, Xu&Cohen et al.[9] built an experimental dataset[1] containing Twitter data and historical stock price data, and proposed a stock movement prediction model called StockNet based on the dataset. This model is a deep generation model, which introduces recursive and continuous latent variables through a new decoder with variational architecture to handle random data, and uses text and price features to predict the stock price movement. Sawhney et al.[4] further construct a relationship graph between each stock through Wikipedia based on the dataset of Xu&Cohen et al., and integrated Twitter data and historical stock price data to obtain the feature representation of each stock using Graph ATtention network(GAT) to predict the movement of each stock.

Although the above mentioned research methods have achieved competitive results, future information is unavoidably used in modeling the feature representations of text and stock prices. Specifically, most researchers derive text and stock price features from global information, which uses not only past social information and historical stock prices, but also future social comments and future stock prices to predict stock price movement on that day. In the real world, however, we cannot get information about the future. Therefore, the features of text and stock price were obtained in this way not only contain a lot of noise, but also lead to the model being over-fitting. In addition, most researchers have modeled social texts and stock prices, separately, ignoring the interaction and correlation between them.

To alleviate the above mentioned problems, in this paper, we propose a method for stock movement prediction based on Temporal Convolution Network (TCN)[13] and Interactive Attention Network (IAN)[14]. In the first, we employed TCN and temporal attention mechanism to obtain the feature representations of Twitter text and stock prices, respectively. And then, two interactive features of Twitter text and stock prices were obtained by using IAN based on the feature representation obtained in the previous step, respectively. Finally, we concatenate two features and use a three-layer neural network to classify for stock movement prediction. The computational framework is shown in Figure 1. In summary, contributions of our study are listed as follows:

- We developed a novel framework for stock movement prediction and achieved competitive results on the experimental dataset.
- We employed TCN and IAN to obtain sufficient feature representations of Twitter text and stock price.
- We regarded the stock movement prediction task as a binary classification problem, and use a three-layer neural network to predict the final results.

II. METHODOLOGY

A. Problem Formulation

In this study, the stock movement prediction task is formalized as a binary classification problem. Therefore, the purpose of this study is to predict the binary movement y of a given stock on the trading day T. where 1 denotes go up and 0 denotes go down:

$$y = \begin{cases} 1, & p_T^c \geq p_{T-1}^c \\ 0, & p_T^c < p_{T-1}^c \end{cases} \tag{1}$$

Where p_T^c denotes the closing price of target stock on trading day T and p_{T-1}^c denotes the closing price of the previous day. In this paper, we expect to use Twitter data and stock price data from several trading days in the training set to predict stock movement in the test set.

B. Temporal Convolutional Network(TCN)

In this study, we chose TCN to extract the feature representations of Twitter data and stock price data, respectively. Compared with traditional CNN and LSTM models, TCN not only can effectively obtain the time series characteristics of text and stock prices, but also can avoid future information being used by models in the process. In Bai et al.[13] and Deng et al.[15] studies, TCN performed significantly better than CNN and Long-Short Term Memory network (LSTM) in both time series processing and stock movement prediction tasks. TCN consists of two main modules, including causal convolution and dilated convolution, which is briefly outlined as follows.

Causal Convolution TCN has two advantages in dealing with time series problems: (i)the output sequence length is equal to the input; (ii)the data will not leak from the future to the past. The original TCN uses the 1D Full-Convolutional Network (FCN) architecture[16] and adds zero padding to keep the sequence length of subsequent layers consistent with that of previous layers. At the same time, TCN uses causal convolution, a strict time constraint model, that is, for the value at time *t* of the previous layer, it only depends on the value at time *t* of the next layer and the value before it. Thus, the model can avoid using future information when obtaining features.

However, the original TCN requires a very deep network or a large-scale filter. Therefore the dilated convolution and residual connection are integrated into the original TCN, so that the final TCN model has long-distance dependent and fast speed in training deep network.

Dilated Convolution Given a one-dimensional input sequence $X \in \Re^n$, the dilated convolution operation F with a filter f on element s of the sequence is denoted as:

$$F(s) = (X *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i} \tag{2}$$

Where d denotes the dilation factor, k denotes filter size, and the $X_{s-d \cdot i}$ denotes the direction of the past. Different from traditional convolution, the sampling rate of dilated convolution is constrained by dilation factor d. For example, d =1 at the bottom layer means that each element is sampled during input, and d =2 at the middle layer means that every two elements are sampled. Therefore, the higher the level, the larger the size of d. In addition, the residual connection has proved to be an effective method for training deep networks[17]. In the final TCN model, a residual block containing two layers of causal dilated convolution and a nonlinear activation function is constructed to replace one layer convolution, and WeightNorm and Dropout are added to each layer to regularize the network:

$$TCN(\bullet) = ReLU(X + F(X)) \tag{3}$$

where X denotes the input block, $F(X)$ denotes causal dilated convolution operation, and $ReLU$ is a nonlinear activation function.

C. Temporal Features of Tweets and Prices

In this part, we first use Glove to obtain the embedding of Twitter text on each trading day and the historical stock price data provided by Feng et al.[10], which are represented by h_i and h_j respectively. Second, h_i and h_j are fed into TCN to obtain the feature representations of h_i and h_j:

$$h_i = TCN(X_{tweet}),\ i \in [1, K] \tag{4}$$

[1]The dataset is available on https://github.com/yumoxu/stocknet-dataset

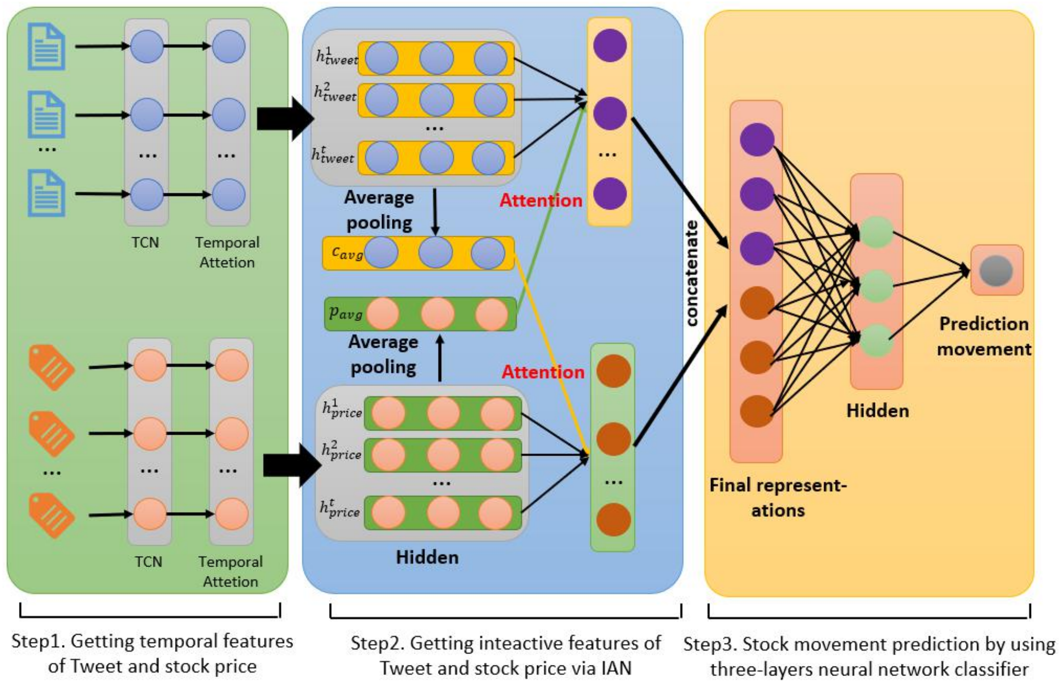

Figure 1. The computational framework for stock movement prediction

$$h_j = TCN(X_{price}),\ j \in [1,T] \tag{5}$$

Where K denotes the number of relevant Twitter opinions of each stock on the i-th trading day. T denotes the current trading day.

As Twitter data and stock prices have different contributions to the prediction of the final movement on different trading days, we use the attention mechanism to obtain the impact weight of different trading days. Finally, the final temporal features of the Twitter data and historical stock price data is obtained, respectively, and is denoted as follows:

$$\alpha_i = softmax(h_i^T tanh\,(Wh_i + b)) \tag{6}$$

$$\beta_j = softmax(h_j^T tanh\,(Wh_j + b)) \tag{7}$$

$$h_{tweet} = \sum_k \alpha_i\, h_i \tag{8}$$

$$h_{price} = \sum_T \beta_j\, h_j \tag{9}$$

where α_i and β_j denotes the attention weight of tweet and price for each trading day. h_{tweet} and h_{price} denotes the temporal feature representations of Twitter opinion and historical stock price.

D. Interactive Features Based on IAN

In order to capture the interactive information between Tweets data and historical stock prices, the Interactive Attention Network (IAN)[14] is employed to obtain the interactive features of text and price. Firstly, we obtain the initial feature representations of tweet and historical price (i.e. c_{avg} and p_{avg}) on all trading days t by averaging the temporal feature:

$$c_{avg} = \sum_{i=1}^{t} \frac{h^i_{tweet}}{k} \tag{10}$$

$$p_{avg} = \sum_{j=1}^{t} \frac{h^j_{price}}{T} \tag{11}$$

Second, considering the impact of social text information on stock prices, we calculate the attention weight γ_i between text representation $[h^1_{tweet}, h^2_{tweet}, \cdots, h^i_{tweet}]$ and the price by:

$$\gamma_i = \frac{exp\,(\sigma(h^i_{tweet}, p_{avg}))}{\sum_{i=1}^{t} exp\,(\sigma(h^i_{tweet}, p_{avg}))} \tag{12}$$

where $\sigma(\bullet)$ denotes the score function for computing the importance of h^i_{tweet} in the tweet data, and is defined as:

$$\sigma(h^i_{tweet}, p_{avg}) = tanh\,(h^i_{tweet} \bullet W_a \bullet p^T_{avg} + b_a) \tag{13}$$

where W_a denotes weight matrix and b_a denotes bias.

Finally, we can obtain the final feature representations C_{tweets} of Twitter text based on attention weight by:

$$C_{tweets} = \sum_{i=1}^{t} \gamma_i\, h^i_{tweet} \tag{14}$$

Similar to the feature representation of Twitter text, we can get the feature representation P_{prices} of historical stock prices affected by Twitter data.

E. *Stock Movement Prediction*

In this study, the stock movement prediction task is regarded as a binary classification problem. we get the final feature H by concatenating the C_{tweets} and P_{prices}. And then, we employed a three-layer neural network classifier to predict stock movement score as follows:

$$\hat{y} = sigmoid(W^l H^l + b^l) \tag{15}$$

Where $H^0 = concat(C_{tweets}, P_{prices})$ denotes the initialized input to the classifier. $sigmoid()$ is an activation function to ensure the movement score is between 0 and 1. W^l and H^l are the weight parameters and bias parameters of the classifier. In addition, we use binary cross entropy as the loss function of the model in the training process.

III. EXPERIMENTS

A. *Benchmark Dataset*

In our study, the benchmark dataset based on the paper of Xu and Cohen et al.[9] and Feng et al.[10]. The dataset of StockNet was built by Xu and Cohen et al, which contains the Twitter dataset and historical stock price dataset. However, we employed the price dataset proposed by Feng et al.[1], due to it containing more price information. In the final experimental dataset, we divided temporally 26614 samples into training set, development set, and test set, as shown in Table 1.

TABLE I. TABLE TYPE STYLES

Type	*Temporal period*	*Sample size*
Training	01/01/2014--01/08/2015	20339
Development	01/08/2015--01/10/2015	2555
Test	01/10/2014--01/01/2016	3720

B. *Experiment Settings*

We employed PyTorh v1.11.0, a Python deep learning framework based on the code of Stocknet[2] and TCN[3] from github. Similar to the parameter settings of these two codes, we set the maximum historical trading day is 5. The maximum message size is set to 30 and the words in the message are set to 40. We obtain the initial word embedding by using the 50-dimensional Glove vectors[18] and train the TCN model with Adam optimizer. In the TCN, the dilation factor d =2, the learning rate is set to 0.0002, and the dorpout is set to 0.5 to avoid over-fitting. In the final three-layer nerual network classifier, the dimension of frist hidden layer is set to 64 and second hidden layer is set to 16. Finally, we evaluate the prediction performance by using accuracy, F1-score, and Matthews Correlation Coefficient (MCC).

C. *Comparison with other Methods*

To evaluate the performance of our proposed method, we selected several methods for comparison and list the comparison results are listed in Table 2.

ARIMA[19]: A method only uses price signals based on technical analysis named AutoRgressive Integrated Moving Average.

TSLDA[20]: A Topic Sentiment Latent Dirichlet Allocation (TSLDA) method, which can capture speech topics and emotions to proves that adding emotional factors can effectively predict stock movement.

HAN[1]: A Hybrid Attention Network (HAN) method includes news-level attention and temporal attention mechanisms and the news data is considered to use for stock movement prediction.

StockNet[9]: A deep learning model that uses historical stock price data and text data of Twitter social platform to predict the stock movement, and has achieved good results in the experiment.

TABLE II. COMPARISON RESULTS WITH OTHER METHODS

Methods	*Accuracy (%)*	*F1-score (%)*	*MCC*
ARIMA[19]	51.4	51.3	-0.0205
TSLDA[20]	54.1	53.9	0.0654
HAN[1]	57.6	57.2	0.0518
StockNet[9]	58.2	57.5	0.0807
ours	**60.4**	**59.4**	**0.2174**

D. *Ablation Study*

In order to explore the impact of each variable module of the model on the overall performance, we conducted an ablation study with three variants. Among them, to find out which factor of Twitter text and historical stock price has a greater impact on the final stock movement prediction, we constructed two variables, including without tweets data and historical stock price data. In addition, To prove that the interaction of Twitter information and historical stock price information can effectively improve the prediction performance, we developed a method without IAN based on our proposed model. Specifically, the three variants of our method are defined as:

w/o Tweets: It only employs the historical stock price data as input for stock movement prediction.

w/o prices: It only employs Twitter text data as input features for stock movement prediction.

w/o IAN: It uses the concatenate features of Twitter text and stock price as the final features and is directly fed into the three-layer neural network classifier, instead of using the interactive attention mechanism to obtain the interactive features of text and price.

The results are shown in Figure 2. The method only employs the Twitter text data as input features have higher values of accuracy and F1-score than the method only uses historical stock price data. Therefore, it proved that adding the text data of social platform can effectively improve model performance. In addition, the method without IAN is slightly lower than our proposed method on the accuracy value and F1-score, so we believe that interactive attention of text and stock price data will help to increase the effectiveness of features for predicting stock movement.

[1]The dataset is available on https://github.com/fulifeng/Adv-ALSTM/tree/master/data/stocknet-dataset/price/ourpped
[2]The code is available on https://github.com/LifangD/Pytorch-Stocknet
[3]The code is available on https://github.com/locuslab/TCN

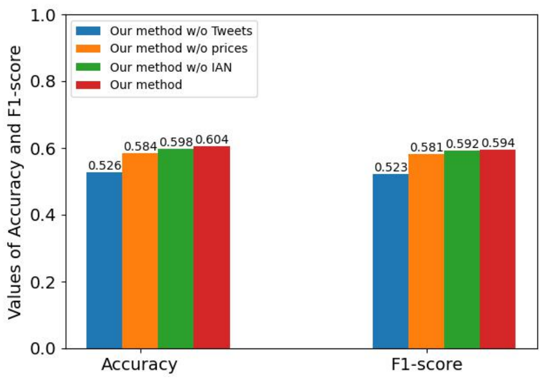

Figure 2. The comparison results between our method and its variants.

IV. Conclusion

In this paper, we developed a novel method based on Temporal Convolution Network (TCN) and Interactive Attention Network (IAN), which used text data of Twitter social and historical stock price data with 11 features[10] for stock movement prediction. In order to alleviate the problem that future information is inevitably used to extract temporal text and historical price features, we use TCN model and temporal attention to obtain effective feature representation of text and price. In addition, considering the interaction between social texts and stock prices, the IAN was introduced to obtain the interactive features between the two factors. Finally, we regarded the stock movement prediction as a binary classification problem, and use a three-layer neural network classifier to get the final prediction movement category.

In our experiments, we first compare our method with several existing methods, and the results show that our method can achieve a competitive performance in stock movement prediction. Second, we compare the impact of each module in our proposed model on the final prediction performance via ablation study. Finally, we found that adding text data from social platforms and interacting with stock prices can significantly improve the prediction performance. However, our method still needs further improvement due to the complexity of the stock market. For instance, as Graph Neural Network (GNN) is more and more popular in dealing with non-structured data, some researchers begin to introduce the correlation between stocks, the knowledge of financial news, and other graph structured data into stock movement prediction[21]. Therefore, how to effectively use GNN and Knowledge Graph (KG) to predict stock movement is one of our future research directions.

References

[1] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proceedings of the eleventh ACM international conference on web search and data mining*, 2018, pp. 261-269.

[2] B. G. Malkiel, *A random walk down Wall Street: including a life-cycle guide to personal investing*. WW Norton & Company, 1999.

[3] I. R. Renu and P. Christie, "Fundamental analysis versus technical analysis–a comparative review," *International Journal of Recent Scientific Research,* vol. 9, no. 1, pp. 23009-23013, 2018.

[4] R. Sawhney, S. Agarwal, A. Wadhwa, and 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 (EMNLP)*, 2020, pp. 8415-8426.

[5] J. Kalyani, P. Bharathi, and P. Jyothi, "Stock trend prediction using news sentiment analysis," *arXiv preprint arXiv:1607.01958,* 2016.

[6] H. D. Huynh, L. M. Dang, and D. Duong, "A new model for stock price movements prediction using deep neural network," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, 2017, pp. 57-62.

[7] 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.

[8] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of Twitter sentiment on stock price returns," *PloS one,* vol. 10, no. 9, p. e0138441, 2015.

[9] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1970-1979.

[10] F. Feng, H. Chen, X. He, J. Ding, M. Sun, and T.-S. Chua, "Enhancing stock movement prediction with adversarial training," *arXiv preprint arXiv:1810.09936,* 2018.

[11] W. Long, L. Song, and Y. Tian, "A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity," *Expert Systems with Applications,* vol. 118, pp. 411-424, 2019.

[12] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[13] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271,* 2018.

[14] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," *arXiv preprint arXiv:1709.00893,* 2017.

[15] S. Deng, N. Zhang, Z. Wen, J. Chen, and H. Chen, "Knowledge-Driven Stock Trend Prediction and Explanation via Temporal Convolutional Network," in *KGTA-WWW-19*, 2019.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532-1543.

[19] R. G. Brown, *Smoothing, forecasting and prediction of discrete time series*. Courier Corporation, 2004.

[20] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1354-1364.

[21] S. Li, J. Wu, X. Jiang, and K. Xu, "Chart GCN: Learning chart information with a graph convolutional network for stock movement prediction," *Knowledge-based systems,* no. Jul.19, p. 248, 2022.