Multi-model based Attention Mechanism for Stock Movement Prediction

Po-Wei Chen[1*], Hao Yuan[2], Jiaying Huang[3], Po-Ju Chen[4]

[1]*School of Mathematics, Sun Yat-sen University, China*
[2]*College of Business Administration, Hunan University, China*
[3]*Guangdong University of Foreign Studies School of Business, China*
[4]*School of Economics, Jinan University, China*
[*]*qq1041604820*

Abstract

Rather than a pure random walk, the stock price changes in the manner of piecewise trend fluctuations. Predictions of stock's future movements have traditionally been based on prior trade data. With the rise of social media, many market participants are opting to make their tactics public, offering a window into the overall's attitude toward future developments by recovering the semantics underlying social media. Social media, on the other hand, includes contradictory information and cannot totally supplant the historical record. In this paper, we present a multimodal attention network that integrates semantic and numerical data to anticipate future stock movements and reduces conflict. We collect semantic information via social media and assess its reliability based on the publisher's name and public reputation. We next design trading strategies by combining semantics from online discussions with numerical characteristics from historical records. The results of our experiments reveal that our strategy exceeds earlier methods in terms of prediction accuracy and trading profit. It demonstrates that our strategy enhances stock movement forecasting performance and guides future multimodal fusion stock forecasting research.

Keywords: *stock movement prediction, attention mechanism, deep learning, forecasting*

1. Introduction

Information influences stock price changes. Information has historically affected the flow of news by disclosing some of the stock market participants' expectations for future movements [1-5]. To anticipate price trends, early studies employed time series of past trend features. Random stock prices, on the other hand, always make projections difficult. Today's social media trends have considerably sped the dissemination of news and given a means of gauging public sentiment toward a certain financial asset. Traditionally, traditional works rely on information such as news and feature engineering to predict its influence on future pricing. Stock forecasting has been researched as an application for NLP downstream jobs, thanks to the prominence of deep neural networks. Researchers have recently used novel learning methodologies and model structures in their study. To better handle unpredictability in markets, Xu et al. [6] utilize neural variational inference. Feng et al. [7] utilize an adversarial training technique to improve prediction accuracy. Liu et al. [8] utilize an attention mechanism and a capsule network to extract richer meaningful sequences from social discourse. On the one hand, the social media ecosystem provides fresh sources for prediction models that go beyond historical data. The variety of the social media landscape, on the other hand, poses some unique issues.

In this research, we offer a new stock's future movement prediction model that combines text information collected on finance forums with historical trend data via the InterIntra attention mechanism to mimic text and prices. In this method, we may deduce the general market perception of the underlying asset based on the aggregated data. On the basis of household-collected data, we assess our system just on stock movement prediction and show that it achieves SOTA performance. We contribute to the creation of a large dataset that includes both time series and text, and we present a new model for analyzing stock movements that incorporates reliability estimates and multimodal processes.

© The Author(s) 2023
G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 1278–1282, 2023.
https://doi.org/10.2991/978-94-6463-042-8_183

2. Related Work

Stock prediction forecasting has been researched for years because of its importance in maximizing stock investing returns. The original strategy relied heavily on historical stock price time series analysis. However, because of the extreme volatility in stock values, forecasting accuracy is restricted. Financial websites now have a vast quantity of available text resources thanks to the advent of the Internet and machine learning, and NLP approaches have become commonplace in this field. Wu et al. [9] and Xu et al. [6] use Twitter to create a corpus. By gathering themes and sentiment on social media, Si et al. developed a topic-based model to forecast stock price fluctuations using sentiment. Liu et al. [8] used Transformer and Capsule networks to mine text semantics in depth and infer actions from them. However, these efforts are impeded by competing meanings and the incomplete nature of a particular social media medium. Some academics have recently used innovative features and learning methodologies to overcome the market's unpredictability. Feng et al. [7] presented an adversarial training strategy to simulate the randomness of price variables by adding perturbations. To enhance information and improve accuracy, Liu et al. [10] used a hierarchical network to fuse financial news and tweets. Despite being a fundamental process, these techniques lack defined metrics to eliminate social media clutter and maximize multimodal information. To get around the simplicity of a single model, we use social influence features and historical trend features. We measure the believability of social media messages using an attention method and integrate their semantic with objective historical trending characteristics. It considerably enhances the stock movement prediction task's performance.

3. Methodology

3.1. Problem Formulation

Our objective is to forecast a stock's movement based on market data at a specific time t . These motions take place between t and $t+\Delta t$, which is the goal time window. The market data we utilize is divided into three sections: Historical pattern data D_t based on daily prices, social media corpus U_t gathered from financial websites, and social influence characteristics. Incorporating poster features X_t and reader comments is a good idea. The size of the market information period is $[t-n+1, t]$, wherein n is the size.

$$Y_t = f(D_t, U_t, X_t) \tag{1}$$

where f is our proposed model, and Y_t is the predictions. Based on previous work, the ground truth for stock movements is regarded as binary labels. Due to market uncertainty, we saw that several goals had a relatively low rate of change. Setting higher and lower stock price change criteria and removing data between them was a frequent technique in prior stock trend prediction projects.

3.2. Embedding Element

Three patterns are embedded using the Embedding Element. There is a historical pattern feature D_t and a social impact feature X_t for each social media post U_t . As a result, the time source H_t is defined as follows:

$$H_t = [D_t, U_t, X_t] \tag{2}$$

In our tests, we discovered that the summing procedure can result in data loss. To maintain rich information, we substitute it with a flatten process. We see the social influence feature as a high-dimensional feature that retains appropriate data and is compatible with textual feature fusion.

$$D^s = DAN(D_t) \tag{3}$$

$$X^e = Linear(X_t) \tag{4}$$

DAN is for deep averaging network, and Linear stands for completely connected layer. A three-dimensional feature map is used to illustrate each historical trend feature. To fully extract spatial information, we use 3D-CNN architecture.

$$U = CNN(U_t) \tag{5}$$

3.3. Fusion Element

In general, investors seek thorough counsel from historical patterns and social media before investing, implying that they are not totally equal but complimentary. As a result, we believe that historical trend data and social media texts complement each other. We define this complimentary impact as a flow of information but update the two methods through an attention process, rather than simply concatenating them. Each characteristic in the same modality is not independently of each other due to the Markov nature of past pricing and frequent interactions on the Internet, therefore there is also an internal flow of information. As a result, in the update schema, we use an internal technique to express the information flow.

The unidirectional complements among past patterns and textual data are computed by the Inter attention block. We start by converting each historical trend data and text data into a query, key, and value, just like the encoder module. The transfer ratio between both the two modes is then calculated:

$$U^{itv} = soft\max(\frac{U_q D_k^T}{\sqrt{d}})D_v$$
$$D^{itv} = soft\max(\frac{D_q U_k^T}{\sqrt{d}})U_v \tag{6}$$

By projection the concatenated of the previous value and translating them into d-dimensional values, we are able to update both schemas:

$$U^{itd} = Linear([H, H^{itv}]^T)$$
$$U^{itd} = Linear([U, U^{itv}]^T) \tag{7}$$

Intra Attention Mechanism obtains a thorough grasp of each update modality by taking into consideration internal impacts. We use an attention mechanism, such as attention blocks, to update values depending on changed keys and queries.

$$U^{inu} = softmax(\frac{U_q^{itd} U_k^{itdT}}{\sqrt{d}})U_v^{itd}$$
$$U^{ind} = Linear(U^{itd} + U^{inu})$$
$$D^{inu} = softmax(\frac{D_q^{itd} D_k^{itdT}}{\sqrt{d}})D_v^{itd} \tag{8}$$
$$D^{ind} = Linear(D^{itd} + D^{inu})$$

Finally, we calculate the average element production among the sequence dimension by feature mapping mechanism.

$$Y = Pooling(H^{ind} \odot C^{ind}) \tag{9}$$

4. Experiments

4.1. Experiment Setting and Data Description

We set the period of the historical pattern description to 32 days and restrict the text samples to 17 days. A batch of 32 shuffled samples has a 7-day prediction interval. The maximum text length is 32 characters. We use an SGD optimizer with a learning rate of 0.005 and a linear decay strategy to train the model. For regularization, we use a 0.1 input dropout rate and a 0.005 weight decay rate. We chose two SOTAs as baselines to test the performances of the proposed model. MHACN and CapTE are the acronyms for MHACN and CapT, respectively.

Our dataset consists of three parts: a social media corpus, social influence descriptions, and historical trend data. The social media corpus was extracted from Snowball, a popular financial communication platform. Considering the timeliness, we collect the texts published from t-l+1 to t, and limit the maximum number of n texts to 96. Since this forum tends to discuss more heavily traded stocks, we selected the top 150 stocks based on their popularity, ensuring a relatively adequate corpus. Historical trend data includes daily open, close, high, low and volume. We calculate and exploit the dispersion between daily highs and lows, and between daily open and close. The number of followers, followers, and posts of the poster describe the basic information of the poster, while his following shares reflect his relevance to topic shares. The poster's trading performance, such as profits from investing in stocks, reflects the credibility of his posts. Readers reply, like, and retweet posts, and we count these actions as numerical features by number of times. We integrate all social influence features into one vector.

4.2. Results Evaluation

Based on Table 1, we conclude that CapT has the best score over the baselines. On the same dataset, the proposed model achieves better accuracy than CapT by more than 3.34%, indicating that our model significantly outperforms the baseline in both accuracy and MCC. We discuss in detail the need to capture textual reliability and our method to complementing semantics with reliability. It shows the complementary effect of social influence features on text features, reducing the conflict of single text modalities. As mentioned above, we assume that historical trend features and text features are complementary. We compared their combined effects with individual effects. We find that our model beats both baselines, suggesting that the two modalities complement each other and the fused features are more predictive. Compared to each partial model mentioned above, the performance of the full model is much better, implying that our hypothesis is robust and multimodal validity.

Table 1. The experimental results

Model	ACC	MCC
MHACN	0.6235	0.0815
CapT	0.6544	0.1026
Ours	**0.6878**	**0.1354**

5. Conclusion

Maximizing use of stock market data can help to improve prediction accuracy. Through capturing multi modalities, we demonstrate a method for predicting stock movements predicated on enriched market information. Our proposed method significantly improves performance both in classification and virtual trading, as shown in the results, introducing a new direction for improving performance: adopt multimodality. Because the market is so complicated, finding enough modalities can help to better show the actual situation. Our research also raises the question of how to perform a successful fusion of modalities in order to improve performance. We will continue to look into relevant features and high-efficiency modality fusion approaches in the future.

References

[1] Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24), 9603-9611.

[2] Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, 164, 163-173.

[3] Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. Expert Systems with Applications, 79, 153-163.

[4] Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. Journal of economic Perspectives, 29(2), 213-38.

[5] Fosso Wamba, S., Kala Kamdjoug, J. R., Epie Bawack, R., & Keogh, J. G. (2020). Bitcoin, Blockchain and Fintech: a systematic review and case studies in the supply chain. Production Planning & Control, 31(2-3), 115-142.

[6] Xu, Y., & Cohen, S. B. (2018, July). 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) (pp. 1970-1979).

[7] Feng, F., Chen, H., He, X., Ding, J., Sun, M., & Chua, T. S. (2018). Enhancing stock movement prediction with adversarial training. arXiv preprint arXiv:1810.09936.

[8] Liu, J., Lin, H., Liu, X., Xu, B., Ren, Y., Diao, Y., & Yang, L. (2019, August). Transformer-based capsule network for stock movement prediction. In Proceedings of the First Workshop on Financial Technology and Natural Language Processing (pp. 66-73).

[9] Wu, H., Zhang, W., Shen, W., & Wang, J. (2018, October). Hybrid deep sequential modeling for social text-driven stock prediction. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 1627-1630).

[10] Liu, J., Lin, H., Yang, L., Xu, B., & Wen, D. (2020). Multi-element hierarchical attention capsule network for stock prediction. IEEE Access, 8, 143114-143123.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.