Adversarial training and attention gated recurrent unit in stock movement prediction

Xi, Li
Supervised by: Yan, Ge

July 3, 2023

Abstract

All investors in the financial market want to have a more accuracy prediction on whether the stock price will increase or decrease. While predicting stock price movements accurately is a challenging task due to the complex and dynamic nature of financial markets. Traditional approaches often struggle to capture the intricate patterns and dependencies in stock data, leading to suboptimal predictions. In recent years, many deep learning tools have shown promising results in various domains. Many researchers have applied deep learning algorithms to financial area. Convolutional Neural Network(CNN) is firstly used in image recognition domain. Adversarial training and attention mechanism are popular deep learning models in the natural language processing and image recognition area. Some researchers have investigated the application of these algorithms in financial sector, like Adv-ALSTM and CNN-LSTM, but there is no research about the combination of CNN and Adv-ALSTM. To save computer resource, we also use Gated Recurrent Unit(GRU) as a replacement of LSTM.

Expected contributions can be summarised as follows:

- Propose a new model which combines CNN with adversarial attentive gated recurrent unit(CNN-AAGRU) and applied it to stock movement prediction.
- Compare the performance of proposed model with Adv-ALSTM, ALSTM, CNN and so on.
- Analyse the results and prove the proposed model outperformed than Adv-ALSTM, ALSTM, CNN.

Ethics statement: This project fits within the scope of the blanket ethics application, as reviewed by my supervisor <Yan Ge>.

I have completed the ethics test on Blackboard. My score is <12>/12.

1 Project Plan

Predicting stock price movements accurately is of paramount importance in financial markets. Investors, traders, and financial institutions rely on accurate forecasts to make informed decisions, optimize investment strategies, and manage risks effectively. The ability to anticipate whether stock prices will rise or fall can lead to significant financial gains or prevent substantial losses. However, the stock market is a complex and dynamic system influenced by a multitude of factors, including economic indicators, company performance, market sentiment, and global events. Traditional methods of stock price prediction, such as technical analysis and fundamental analysis, have limitations in capturing the intricate patterns and non-linear relationships within the data[1]. Therefore, the development of advanced predictive models using machine learning and deep learning techniques has gained considerable attention in recent years.

However, despite the extensive research conducted in this field, achieving high prediction accuracy remains a challenge. Many existing models have yielded suboptimal results, with accuracies comparable to random guessing. They often struggle to capture the complex dependencies and temporal dynamics inherent in stock market data. This limitation can be attributed to the inability of conventional models to effectively extract and utilize informative features from the data. For instance, in [2], the achieved accuracy was reported to be only 57%, slightly better than flipping a coin. This emphasizes the need for more precise and reliable prediction models.

1.1 Motivation

Motivated by the limitations of existing approaches, this research aims to bridge the research gap by proposing a novel model that combines the strengths of convolutional neural networks (CNNs) and the adversarial attention long short-term memory (Adv-ALSTM) model. By integrating these two powerful architectures, we seek to enhance the accuracy and robustness of stock price predictions. The CNN component of the proposed model excels in feature extraction and pattern recognition, allowing it to capture relevant spatial and temporal dependencies within the input data. Meanwhile, the Adv-ALSTM model's attention mechanism enables it to focus on informative aspects of the input sequence, enhancing the model's ability to identify critical patterns and make accurate predictions.

By leveraging the complementary strengths of CNN and Adv-ALSTM, our proposed model aims to provide more precise and reliable predictions of stock price movements. We anticipate that this integrated approach will outperform existing models and yield significant advancements in the field of stock price prediction.

1.2 Proposed Solution

The proposed model combines the feature extraction capabilities of CNNs with the attention-based mechanism of Adv-ALSTM to enhance the prediction accuracy of stock price movements. Firstly, the CNN component receives input data, and performs convolutional operations to extract relevant features. This enables the model to identify significant spatial dependencies within the data. Subsequently, the output of the CNN component is fed into the Adv-ALSTM model, which incorporates attention mechanisms to focus on the most informative features and capture long-term dependencies. The attention mechanism allows the model to dynamically weigh the importance of different input features and allocate more attention to crucial patterns within the sequence.

By combining the strengths of CNNs and Adv-ALSTM, our proposed model aims to overcome the limitations of existing approaches, achieve more accurate stock price predictions and have higher computational efficiency. We expect that the integration of these two powerful architectures will enhance the model's ability to capture complex relationships within the data, leading to improved forecasting performance.

1.3 Structure

The dissertation is organized as the following structure. Chapter 1 is the introduction, highlights the research's importance, presents the research gap, and outlines the proposed model. Chapter 2 is background, presents a comprehensive review of related works, discussing existing approaches, their strengths and limitations, and identifying the research gap. Chapter 3 talks about the project execution, including data, model architecture, and training procedures. Chapter 4 is the critical evaluation, presents the experimental results and performance evaluation of the proposed model, comparing it against baseline models and analyzing its predictive capabilities. Chapter 5 is conclusion, discusses the findings, implications, and limitations of the study, and provides recommendations for future research.

In conclusion, this dissertation aims to contribute to the field of stock price prediction by proposing an innovative model that leverages the synergy between CNNs and Adv-ALSTM.

2 Literature Review

As mentioned in the Project Plan, stock price movement prediction is a challenging and critical task in financial markets. Recent years, the application of deep learning algorithms to predict stock price movements has gained substantial interest due to their ability to capture complex patterns and relationships in financial data.

2.1 Recurrent Neural Network

Recurrent Neural Network(RNN) is a type of deep learning network for time series data. The basic operation of it involves iterating through the input sequence one element at a time. At each time step, it takes an input vector and combines it with the hidden state from the previous time step to generate an output and update the current hidden state. This feedback mechanism enables RNN to capture temporal dependencies and learn patterns in sequential data[3].

2.2 Long Short Term Memory

While standard RNNs suffer from the vanishing gradient problem and limited memory capacity, long short term memory(LSTM) addresses these issues through the introduction of memory cells and gating mechanisms[4]. It can remember both short term and long term values and mostly being used with time series data[5]. The key innovation in LSTM is the incorporation of memory gates, including the input gate, forget gate, and output gate. These gates control the information flow and allow the network to learn which information to remember, discard, or output at each time step [4].

In recent years, there has been significant researches applying LSTM to the stock prediction problem. In 2015, [6] applied LSTM to predict stock price using the data from China stock market. The superiority of the LSTM model in stock returns prediction was verified by comparing it with the traditional linear regression model and the support vector machine (SVM) model. In 2016, [7] applied LSTM to Romanian stock markets to predict the future trend. In 2017, [8] study combined wavelet denoising with LSTM for stock price prediction. The inclusion of wavelet denoising helps filter out small market fluctuations, allowing the model to capture the broader market trends more accurately. This approach outperformed the original LSTM model, but the authors emphasized that real financial markets are more complex, requiring domain experts to analyze the results in more detail for practical applications.

2.3 Gated Recurrent Unit

With LSTM mitigate the vanishing and exploding gradient problems encountered in traditional RNNs, it faces some challenges as having too many parameters as well its long training time[9]. In light of these concerns, a variation of LSTM called Gated Recurrent Unit(GRU) was proposed by [10]. GRU inherits the gate mechanisms from LSTM but combines the forget gate and input gate, thereby simplifying the architecture, reduced parameter count which and can lead to faster training and inference times, making GRU more computationally efficient.

2.4 Attention Mechanism

The attention mechanism was initially proposed in [11] and has gained widespread application in the field of machine learning. It aims to enhance the model's focus on relevant information by assigning different weights to different parts of the input sequence. The weights which indicate the importance of each element in the context of the current prediction are obtained by comparing the similarity between the current hidden state of the LSTM and each element in the input sequence. Attention Mechanism allows the model to selectively attend to important features or time steps, improving its ability to capture dependencies and make predictions.

In 2018, [12] and [7] both investigated the application of LSTM with attention mechanism. [12] using a feed-forward neural network, the previous decoder output and the current encoder hidden state are taken as input, and outputs an alignment score, which indicates the importance of the current encoder hidden state for the decoder output. [12] compares the performance of LSTM and LSTM with attention on forecasting financial time series, using five stocks from Kaggle's Two Sigma dataset. It hypothesizes that attention can help overcome the long-term dependency problem experienced by LSTM models. While in [7] using bilinear function-based attention mechanism. The bilinear function computes the pairwise interactions between the LSTM hidden state and input elements. It applies a matrix multiplication between the LSTM hidden state and a learnable weight matrix, followed by an element-wise multiplication with the input elements. The resulting scores represent the alignment between the LSTM hidden state and each input element.

The main difference between these two approaches lies in their computational mechanisms. The feed-forward neural network in [12] learns a complex mapping from the LSTM hidden state and input elements to alignment scores, allowing for more flexibility in capturing nonlinear relationships. In contrast, the bilinear function in [7] explicitly models pairwise interactions between the LSTM hidden state and input elements, potentially capturing more fine-grained dependencies.

In 2019, [13] using attention LSTM to Hong Kong stock price, the paper also confirmed that ALSTM performed better than the traditional LSTM.

2.5 Adversarial Training

Adversarial training, originally proposed by [14] in 2014, is a machine learning technique that involves training a model in a game-like setting, where two components, namely the generator and the discriminator, are pitted against each other. The basic principle is to optimize the generator to produce outputs that can deceive the discriminator, while simultaneously training the discriminator to accurately distinguish between real and generated samples.

Adversarial training is primarily applied in various domains such as image recognition, natural language processing(NLP) and image generation.

In image recognition, [15, 16] proved that adversarial training enhances models' capacity to accurately identify images under varying conditions such as diverse lighting, angles, and occlusions. Additionally, it serves as a defense mechanism against malicious adversarial attacks aiming to deceive the model. Within the realm of NLP, [17] improves models' comprehension of text across different language styles, grammar errors, and spelling mistakes by incorporating adversarial training. In the domain of image generation, both [18] and [19] reveals that adversarial training enables models to generate high-quality and diverse images. It ensures that the generated images adhere to the real data distribution and avoids generating images with noticeable flaws.

In recent years, adversarial training has also found application in financial domain, as evidenced by the work presented in [2]. This expansion of adversarial training into diverse domains reflects its growing significance and potential for addressing challenges across different fields.

Building upon the attention mechanism LSTM, [2] introduced Adversarial Training Attention LSTM (Adv-ALSTM), which combines the principles of attention mechanism and adversarial training. Adv-ALSTM model extends the traditional LSTM architecture by incorporating an adversarial training component specifically designed for the attention mechanism. In the Adv-ALSTM, the data first goes through a feature mapping layer, then passes through the LSTM layer in the model, and finally undergoes the attention mechanism to generate the prediction layer. To enhance the model's generalization ability, adversarial training is employed. In the adversarial training component, perturbations are added in the direction that has the maximum impact on the experimental results, that is the . Experimental results have demonstrated that Adv-ALSTM outperforms baseline models such as LSTM, ALSTM, and StockNet.

2.6 Convolutional Neural Network

Convolutional Neural Network(CNN) is a wildly used supervised classification model[20]. CNNs employ convolutional filters or kernels that slide over the input data, performing element-wise multiplication and summation operations. By using multiple convolutional filters, CNNs can learn diverse and hierarchical representations of the input data.

[21, 22, 23] investigated the application of CNN on stock price prediction. [21] and [23] using data including the industry information, company financial statements and so on. [22] only using one dimension stock price as data input. [24] firstly combined CNN with LSTM to make quantitative strategy on stock selection task. It used CNN for stock ranker to achieve the quantitative stock selection strategy and used LSTM to achieve the quantitative timing strategy. [25] used LSTM for time series data and used CNN for the graph of each stock, such as line charts, bar charts, K-charts and candle charts. The model combined the output of LSTM and CNN, and got good performance on stock price predicting.

As mentioned above, most researches separately applied CNN and LSTM on stock prediction task. Part of researches thought about the combination, but they did not mention the feature combination. Our work will based on [2], combine CNN with Adv-ALSTM and replace LSTM with GRU to get a more comprehensive feature to better predict the stock movement and have a higher computational efficiency.

References

[1] A. Chatterjee, H. Bhowmick, and J. Sen, "Stock price prediction using time series, econometric, machine learning, and deep learning models," in *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*. IEEE, 2021, pp. 289–296.

[2] 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.

[3] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied soft computing*, vol. 90, p. 106181, 2020.

[6] K. Chen, Y. Zhou, and F. Dai, "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)*. IEEE, 2015, pp. 2823–2824.

[7] L.-C. Cheng, Y.-H. Huang, and M.-E. Wu, "Applied attention-based lstm neural networks in stock prediction," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 4716–4718.

[8] Z. Li and V. Tam, "Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2017, pp. 1–8.

[9] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *CoRR*, vol. abs/1503.04069, 2015. [Online]. Available: http://arxiv.org/abs/1503.04069

[10] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[12] T. Hollis, A. Viscardi, and S. E. Yi, "A comparison of lstms and attention mechanisms for forecasting financial time series," *arXiv preprint arXiv:1812.07699*, 2018.

[13] S. Chen and L. Ge, "Exploring the attention mechanism in lstm-based hong kong stock price movement prediction," *Quantitative Finance*, vol. 19, no. 9, pp. 1507–1515, 2019.

[14] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[15] ——, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[16] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.

[17] T. Miyato, A. M. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," *arXiv preprint arXiv:1605.07725*, 2016.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] L. Di Persio, O. Honchar *et al.*, "Artificial neural networks architectures for stock price prediction: Comparisons and applications," *International journal of circuits, systems and signal processing*, vol. 10, pp. 403–413, 2016.

[22] H. Gunduz, Y. Yaslan, and Z. Cataltepe, "Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations," *Knowledge-Based Systems*, vol. 137, pp. 138–148, 2017.

[23] E. Hoseinzade and S. Haratizadeh, "Cnnpred: Cnn-based stock market prediction using a diverse set of variables," *Expert Systems with Applications*, vol. 129, pp. 273–285, 2019.

[24] S. Liu, C. Zhang, and J. Ma, "Cnn-lstm neural network model for quantitative strategy analysis in stock markets," in *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24*. Springer, 2017, pp. 198–206.

[25] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion lstm-cnn model using different representations of the same data," *PloS one*, vol. 14, no. 2, p. e0212320, 2019.

A Project Timeline

The project plan is submitted on 3rd July, so the timeline begin from 4th July.

- Week 1 (7.4-7.8): On holiday
- Week 2 (7.9-7.16): Finish all the coding work.
- Week 3&4 (7.17-7.30): Design the experiments and record the results.
- Week 5 (7.31-8.6): Visualize the experiments' results and begin to write the Chapter 1, 2.
- Week 6 (8.7-8.13): Write Chapter 3, the project execution part.
- Week 7 (8.14-8.20): Write Chapter 4, conclusion and all the other parts.
- Week 8 (8.20-8.27): Proofreading the draft.
- Week 9 (8.28-8.31): In case some plans are not completed.
- Week 10 (9.1-9.7): Prepare the video presentation.

B Risk Assessment

Table 1: Risks and Mitigations

Risk	Likelihood	Severity	Mitigation
Computer problem (stolen or broken)	Low	High	Backup all work using OneDrive
New model performs worse	Medium	High	Change a model or dataset to mitigate the contingency
Fall ill	Medium	High	Health is more important, keep relax until come back to life
Plan was disrupted	High	High	Making plans early before the submission deadline