Enhancing Stock Price Forecasting and Trading Strategy through Bidirectional LSTM Integration

JIAYI LIU*

University of Utah, David Eccles School of Business, Utah, United States. liujiayi27@163.com

Yufeng Yang

Zhejiang University, Zhejiang, China. 3200102634@zju.edu.cn

Teng Lin

Syracuse University Martin J. Whitman School of Management, New York, United States. tlin07@syr.edu

Chuanhui Peng

Washington university in St louis. United States. chuanhui@wustl.edu

Yancong Deng

University of California, San Diego. Jacob School of Engineering, United States. yancongdeng@gmail.com

Abstract

Accurately predicting data flow is a significant challenge in industrial automation, especially given the diversity of data types. Traditional time series prediction models often struggle to consistently produce effective predictions across varied data sets. To address these limitations, this paper explored time-series prediction model with traditional LSTM and Bidirectional LSTM (Bi-LSTM) architectures. Experimental results indicate that our proposed Bi-LSTM model offers superior prediction accuracy compared to the standalone LSTM algorithm. Furthermore, leveraging our prediction model, we develop an efficient trading strategy algorithm that potentially yields a positive capital gain compared to a holding position.

CCS CONCEPTS • Applied computing • Operations research • Forecasting

Additional Keywords and Phrases: Bidirectional LSTM, Stock Price Forecasting, Trading Strategy Optimization, Active Portfolio Management

1 INTRODUCTION

The intricate and volatile nature of the financial market makes the precise prediction of stock prices a formidable task. Over the years, various traditional time series models, such as the Autoregressive

* Place the footnote text for the author (if applicable) here.

Conditional Heteroscedastic (ARCH) and Generalized Autoregressive Conditional Heteroskedastic (GARCH) models, have been employed extensively to gauge volatility in financial time series data. However, these models often hinge on stationary assumptions, which are difficult to meet when dealing with real-world financial data. This inherent limitation often compromises their efficacy on out-of-sample data.

In pursuit of enhanced precision in stock price forecasting, the research community has pivoted towards artificial intelligence and machine learning methodologies. Within this spectrum, neural network-based techniques, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), have shown commendable efficacy [1-3]. Of particular note, the Long Short-Term Memory (LSTM) algorithm, a variant of RNN, has garnered significant interest. Its prowess in adeptly handling long-term dependencies and intricate sequential data sets it apart.

This paper introduces a hybrid model for stock price forecasting, amalgamating both LSTM and Bidirectional LSTM (Bi-LSTM) architectures. Our objective in blending these models is to refine prediction accuracy and more adeptly discern complex patterns inherent in stock market data. The LSTM model, with its memory and hidden state capabilities, empowers us to distill pivotal temporal insights from historical stock price trajectories [10]. Concurrently, the Bi-LSTM model, with its bidirectional information flow encompassing both past and future observations, offers an enriched understanding of the sequential data.

2 RELATED WORK

Engle (1982) pioneered the introduction of the autoregressive conditional heteroscedastic (ARCH) model to estimate the means and variances of inflation in the United Kingdom [4]. Building upon Engle's foundational work, Andersen and Bollerslev (1986) developed the generalized Autoregressive Conditional Heteroskedastic (GARCH) model, offering a more logically coherent lag structure [5]. Subsequently, Nelson (1991) proposed the EGARCH model to address conditional heteroscedasticity or volatility clustering, in innovation processes [6]. GARCH-type models adeptly captured the nuances of in-sample volatility in financial time series data. The majority of the earlier works in this domain predominantly employed these linear models. Despite their statistical interpretability and widespread application in estimating financial sector volatility, these traditional time series models often grapple with inherent stationary assumptions. These assumptions are frequently misaligned with the dynamic nature of real-world financial data, compromising their efficacy on out-of-sample datasets.

In an endeavor to ameliorate these shortcomings, Andersen and Bollerslev (1998) introduced the concept of realized volatility (RV). Notably, RV is agnostic to the specific assumptions underpinning models used to quantify volatility and mitigates measurement errors by harnessing high-frequency data [7]. Shao and Yin (2008) constructed a realized volatility model and a realized range model to ultimately calculate VaR (Value at Risk) utilizing intraday high-frequency data [8]. Their work underscored that models grounded in intraday data outperformed those relying on daily returns. By leveraging high-frequency data, RV illuminates previously unobservable volatility nuances. In light of this, it's cogent to treat RV, derived from high-frequency data, as a reflection of genuine volatility within this research.

Parallel to these developments, the ascendancy of artificial intelligence and burgeoning computational power catalyzed the integration of potent machine learning techniques into financial time series modeling. Approaches span a gamut from support vector machines (SVM) and Random Forests (RF) [9, 10] to neural network architectures like artificial neural networks (ANN), Convolutional Neural Networks (CNN), and

Recurrent Neural Networks (RNN). Within this neural spectrum, deep architectures like Long Short-Term Memory (LSTM) units have carved a niche. Barunik and Krehlik (2016) blazed a trail by deploying ANNs to forecast volatility in the energy market, amplifying prediction accuracy through high-frequency data [11]. ANNs, intrinsically nonparametric and data-driven, have gained traction with the influx of voluminous financial datasets. Hochreiter and Schmidhuber (1997) heralded the advent of the LSTM algorithm, a prominent RNN variant, adept at tackling intricate, long time-lag tasks [12]. Contrasting traditional predictive models, RNNs accrue memory from data trajectories and unveil hidden state features to decipher statistical patterns, obviating the need for extensive macroeconomic and company-specific variable calibrations. This holistic mapping system offers the luxury of nonparametric statistical inference. LSTMs, in particular, are favored in practice to sidestep long-term dependency challenges. Chen, Zhou, and Dai (2015) showcased the prowess of LSTMs in predicting Chinese stock returns, underscoring their potential in stock market forecasting [13]. In the present study, I endeavor to probe the viability of LSTMs in predicting multiple stock prices.

3 METHODOLOGY

3.1 Long Short-Term Memory

Recurrent Neural Networks (RNNs) are renowned for their prowess in predicting sequential data, encompassing input, hidden, and output layers. However, a cardinal limitation of classical RNNs is the vanishing gradient problem. LSTM, an advanced architecture, was conceived to mitigate this impediment [10]. The feed-forwarding process of LSTM, given input data x_t and hidden state h_t at time-step t can be delineated as follows:

$$i_t = \sigma(\mathrm{W}_1 X + \mathrm{b}_1) \qquad (5)$$

$$\mathrm{f_t} = \sigma(\mathrm{W_2 X + b_2}) \qquad (6)$$

$$\mathrm{o_t} = \sigma(\mathrm{W_3 X + b_3}) \qquad (7)$$

$$\mathrm{g_t} = \tan\mathrm{h}(\mathrm{W_4 X + b_4}) \qquad (8)$$

$$\mathrm{c_t} = \mathrm{c_{(t-1)}} \times \mathrm{f_t} + \mathrm{g_t} \times \mathrm{i_t} \qquad (9)$$

$$\mathrm{h_t} = \tanh(\mathrm{c_t}) \times \mathrm{o_t} \qquad (10)$$

Where W_i and b_i are weights and bias terms, respectively, and $X = \binom{b_t}{h_{t-1}}$. The function σ and tanh are defined by $\sigma = \frac{1}{(1+e^{-x})}$ and $tanh = \frac{e^x + e - x}{e^x - e - x}$.

3.2 Bidirectional LSTM

The Bidirectional RNN represents an enhancement of the traditional RNN, devised to assimilate input data sequences from both past and future contexts. The Bidirectional LSTM architecture deploys two distinct layers: one layer processes the input data sequences in their original order, while the other processes them in reverse. The outcomes from both layers can be amalgamated using various merging strategies. Empirical evidence from numerous applications underscores the superiority of Bidirectional LSTM over its traditional counterpart in certain contexts.

3.3 Hybrid Model

Historical research accentuates that the fusion of neural networks with GARCH-type models can augment prediction accuracy. Extrapolating from this, it's plausible that the synthesis of diverse neural network architectures might yield similar enhancements. The LSTM model has already showcased its efficacy in stock price forecasting. By integrating Bi-LSTM layers into this hybrid construct, the aspiration is to harness its bidirectional capabilities. This inclusion aims to yield a more holistic understanding of sequential data, thereby potentially elevating the model's performance.

4 EXPERIMENT

4.1 Data Selection and Preparation

For the purposes of this study, 11 stocks spanning 8 distinct sectors were selected, as outlined in Table 1. The historical trading data employed for these experiments was sourced from online databases. In training the LSTM model, a holdout method was adopted where 90% of the data constituted the training set, while the remaining 10% served as the validation set for hyperparameter tuning. This diversity is quintessential to examining the robustness of the LSTM models across sectors, each exhibiting unique volatility patterns and market dynamics. The LSTM and Bi-LSTM models are sophisticated in their architecture and ensuring a proper training-validation split is critical to prevent overfitting and to ensure the model generalizes well to unseen data.

Table 1: 11 stocks from 8 different sectors

Symbol	Sector
AA	Industrials
AAPL	Technology
ACT	Finance
AMGN	Health Care
BAX	Health Care
BK	Finance
C	Finance
CAT	Industrials
V	Consumer Discretionary
VZ	Telecommunications
WBA	Consumer Staples

4.2 Single Stock Price prediction

The rolling time window approach was employed for stock price prediction. Figures 1 to 11 delineate the out-of-sample Mean Squared Error (MSE) for each model. A comparative analysis between the LSTM and Bidirectional LSTM networks revealed enhanced performance with the Bidirectional LSTM. This suggests that the Bidirectional LSTM offers a more refined approach for stock price forecasting than its traditional LSTM counterpart. These insights can be instrumental in refining stock price prediction methodologies and in shaping informed investment strategies.

The following images Figure 1 to Figure 11 demonstrate our LSTM and BiLSTM model's performance on individual stock pricing prediction on popular Stock sectors, AA, AAPL, ACT, AMGN, BAX, BK, C, CAT, V, VZ, and WBA respectably.

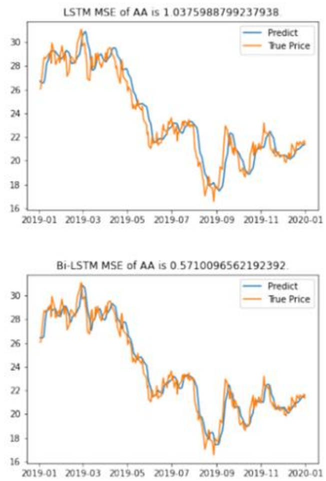

Fig 1: Stock AA Prediction

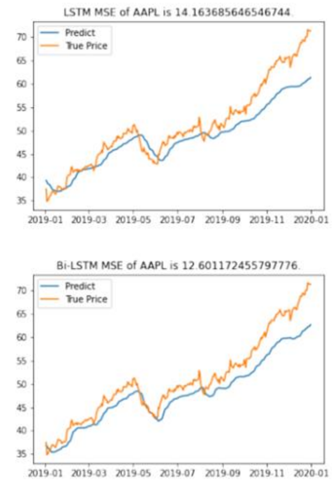

Fig 2: Stock AAPL Prediction

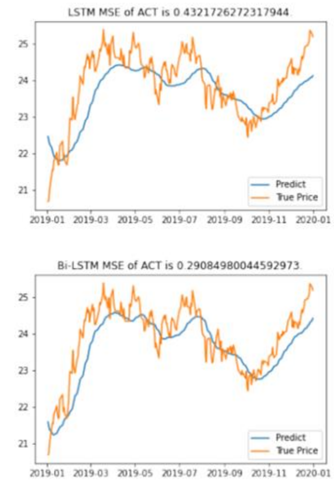

Fig 2: Stock ACT Prediction

Fig 1 – Fig 3: comparison between the prediction performance of the LSTM and Bidirectional LSTM neural networks on Stock AA, AAPL, and ACT

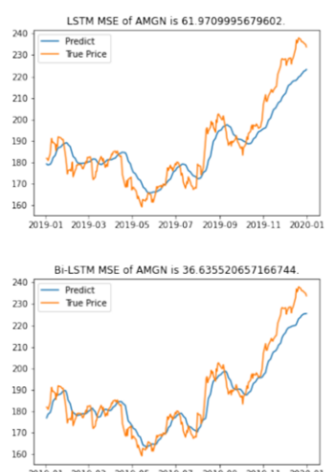

Fig 4: Stock AMGN Prediction

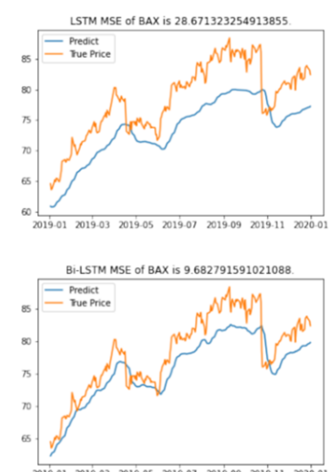

Fig 5: Stock BAX Prediction

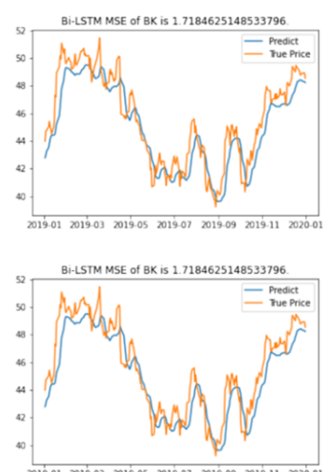

Fig 6: Stock BK Prediction

Fig 4 – Fig 6: comparison between the prediction performance of the LSTM and Bidirectional LSTM neural networks on Stock AMGN, BAX, and BK

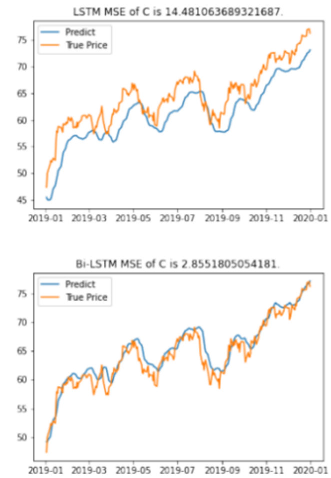

Fig 7: Stock C Prediction

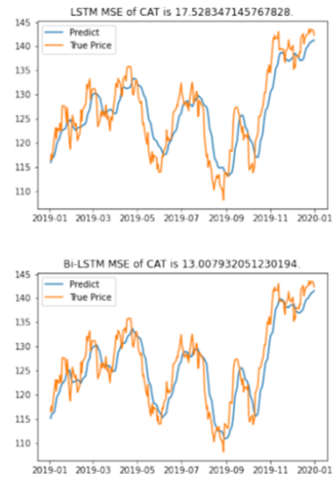

Fig 8: Stock CAT Prediction

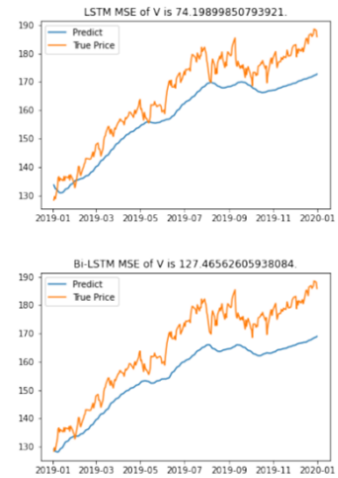

Fig 9: Stock V Prediction

Fig 7 – Fig 9: comparison between the prediction performance of the LSTM and Bidirectional LSTM neural networks on Stock C, CAT, and V

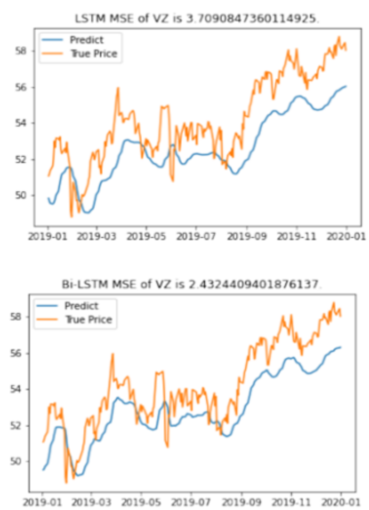

Fig 10: Stock VZ Prediction

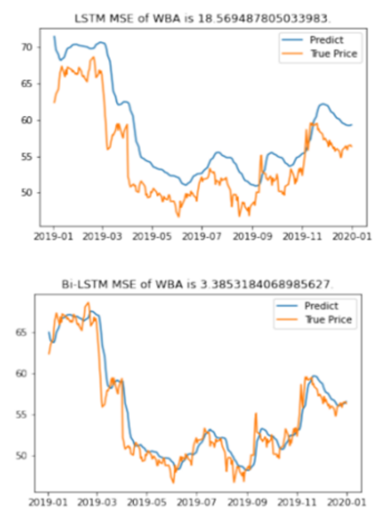

Fig 11: Stock WBA Prediction

Fig 10 – Fig 11: comparison between the prediction performance of the LSTM and Bidirectional LSTM neural networks on Stock VZ, WBA

As a result, from the above figures, they showcase the performance over a one-year prediction horizon, additional tests were conducted to gauge short-term and long-term prediction capabilities. For the short-term analysis, networks were calibrated to forecast stock prices six months into the future. Consistent with the one-year prediction findings, the Bidirectional LSTM exhibited superior accuracy for the majority of stocks compared to the traditional LSTM. For long-term predictions spanning two years, the Bidirectional LSTM once again outperformed the LSTM. The evident superiority of the Bidirectional LSTM over the traditional LSTM underscores the importance of considering the data sequences in time series forecasting.

4.3 Trading Strategy with Portfolio

The formulated trading strategy, though rudimentary, serves as a litmus test for the model's real-world applicability. By defining adjustment frequencies, the strategy acknowledges market realities where frequent trading can incur costs, and infrequent adjustments can miss out on market opportunities.

To evaluate the effectiveness of our proposed trading approach, we designed a basic trading strategy that encompasses the following steps:

- Define adjustment frequencies at intervals of 3, 5, 7, and 14 days.
- Harness the predictive model to periodically identify the top 5 stocks poised to yield the highest returns over the subsequent adjustment window.
- Distribute the investment capital across these stocks, weighted by their projected returns. Specifically, stocks with higher anticipated returns would secure a larger allocation.

Our Proposed model dynamically selects the top 5 stocks based on predicted returns. This process is an embodiment of active portfolio management. While traditional portfolio management often focuses on diversification to minimize risk, this active strategy maximized returns based on predictive insights, a paradigm shift made possible by advanced machine learning models.

Executing this trading strategy enables us to ascertain its merit and potential to outperform a static investment approach. For a holistic evaluation, an initial investment corpus of $100,000 was assumed. The comparative analysis against a static investment approach enables the potential gains achievable through active management through the proposed forecasting model. As the result showed in Figure 12 to Figure 15, they demonstrate how the proposed model performs over different intervals adjustment frequencies. As a result, the improvements observed with the 7-day and 14-day adjustment windows suggest that medium-term trading strategies, as opposed to very short or very long-term, might strike the optimal balance between capturing market opportunities and mitigating transaction costs.

Figure 12: portfolio's performance using 3 days as the adjustment window

Figure 13: portfolio's performance using 5 days as the adjustment window

Figure 14: portfolio's performance using 7 days as the adjustment window

Figure 15: portfolio's performance using 14 days as the adjustment window

While the results are promising from our proposed model, it's imperative to acknowledge that stock markets are influenced by a myriad of factors, including geopolitical events, policy changes, and global economic conditions. Future iterations of this work could consider incorporating external variables or news sentiment analysis to further refine the model's predictive capabilities. Furthermore, real-world implementation should consider transaction costs, tax implications, and other practical aspects that can influence net returns.

5 RESULTS AND CONCLUSION

Our experiment involved the construction of a dynamic trading strategy, the performance of which was benchmarked against a passive holding approach. The results were illuminating. A marked enhancement in portfolio performance was discerned when periodic adjustments were made, especially on a 7-day and 14-day cadence. The figures vividly depict this disparity, with the adjusted portfolios consistently outpacing their static counterparts in terms of final asset values.

This outcome accentuates the merits of active portfolio management, particularly when it is informed by sophisticated predictive models like the LSTM. By leveraging the insights provided by our model and proactively adjusting the portfolio, investors can potentially realize significantly higher returns. The results serve as a testament to the potential of machine learning in reshaping traditional investment strategies, paving the way for a new era of data-driven decision-making in the financial sector.

REFERENCES

[1] AKGÜL, İsmail, and Volkan KAYA. 2022. A review on Artificial Neural Networks.

[2] Aloysius, Neena, and M. Geetha. "A review on deep convolutional neural networks." 2017 international conference on communication and signal processing (ICCSP). IEEE, 2017.Kaur M. and Mohta A. A Review of Deep Learning with Recurrent Neural Network., 2019.

[3] Engle, Robert F. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." Econometrica: Journal of the econometric society (1982): 987-1007. Tim Bollerslev. 1986. Generalized autoregressive conditional heteroskedasticity. J. Econom. (1986), 307-327. DOI https://doi.org/10.1016/0304-4076(86)90063-1

[4] Nelson, Daniel B. "Conditional heteroskedasticity in asset returns: A new approach." Econometrica: Journal of the econometric society (1991): 347-370.

[5] Andersen, Torben G., and Tim Bollerslev. "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts." International economic review (1998): 885-905.

[6] Xi Dong Shao and Lian Qian Yin. 2008. The study on financial market risk measures in China based on realized range and realized volatility. J. Financ. Res. (2008), 109-121.

[7] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20 (1995): 273-297.

[8] Breiman, Leo. "Random forests." Machine learning 45 (2001): 5-32.

[9] Barunik, Jozef, Tomas Krehlik, and Lukas Vacha. "Modeling and forecasting exchange rate volatility in time-frequency domain." European Journal of Operational Research 251.1 (2016): 329-340.

[10] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[11] Chen, Kai, Yi Zhou, and Fangyan Dai. "A LSTM-based method for stock returns prediction: A case study of China stock market." 2015 IEEE international conference on big data (big data). IEEE, 2015.