Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/26/20230573

Application of BiLSTM-Transformer in Portfolio Optimization

Chuchu Sun [1, a,*,†], Haoqin Li [2, †], Sihan Fu [3, †]

[1]*School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

[2]*Jinan University – University of Birmingham Joint Institute, Jinan University, Guangzhou 511443, China*

[3] *College of Science and Technology, Wenzhou-Kean University, Wenzhou 325060, China*

†These authors contributed equally.

a.sccnuaa@nuaa.edu.cn

**corresponding author*

Abstract: Portfolio optimization models that use predictions can effectively capture short-term investment opportunities. However, in traditional models, inaccurate predictions of the expected excess return of different assets can negatively impact investment performance. Deep learning models have demonstrated significant advantages over time series models in this regard. This paper connects Transformer model and the BiLSTM model, which is short for bi-directional Long Short-Term Memory, for return prediction for portfolio model performance enhancement. To be specific, the model of BiLSTM-Transformer is firstly applied for predicting the yield of alternative assets, which is then incorporated in the mean–variance (MV) model. Using 6 component stocks of the US30 index as alternative assets, 270 investments are conducted, and the empirical results are compared with LSTM and Transformer model. The comparison verifies the superiority of BiLSTM-Transformer model in improving prediction accuracy and boost of portfolio model performance.

Keywords: prediction, portfolio, optimization, stock, asset

1. Introduction

To solve the portfolio optimization problem for various asset classes, Markowitz proposed the famous mean-variance model (Mean-Variance, MV), which measures the expected return generated by alternative asset, and the risks caused by the matrix with variance-covariance between the returns of each asset [1-2]. The model is widely applied by investors and researchers and has important theoretical and practical implications. However, the traditional portfolio optimization models commonly rely on historical mean return as the expected return, which may lead to unreliable return estimation in the short run [3]. To address this issue, numerous researchers have proposed to use prediction of return as the expected one to construct optimized portfolio model [3-5].

Traditionally, time series forecast tackling was done through analyzing time series linear method [3]. While due to the influence of different factors like macroeconomics, investor sentiment and government policies, the financial market shows characteristics such as nonlinearity, nonstationary, and

DOI: 10.54254/2754-1169/26/20230573

complexity [6]. Recently, researchers have attempted to use machine learning-based artificial intelligence methods to address problems that traditional statistics cannot solve, providing predictive models for financial market analysis and investment decision-making [7]. In the field of machine learning, deep learning has emerged and showed superior performance especially in predicting financial market trends [8]. Among that, DMLP, which is short for deep multilayer perceptron, and the neural network of LSTM (which is short for Long Short Term Memory) and CNN which is short for convolutional neural network, are commonly used [9].

Therefore, in addition to traditional time series models and machine learning models, some scholars also proposed the integration of deep learning about techniques of return prediction in the process of forming portfolio, thus enhancing its original optimized model performances. For instance, For instance, Ma et al. have explored how to respectively use machine learning models and deep learning, such as the neural network of LSTM and DMLP as well as CNN, which are combined with the portfolio formation with return prediction [10]. Meanwhile, Zhang et al. have utilized a traditional portfolio model with transformer model predicted results to construct an optimal portfolio, and have compared the empirical results with those obtained from the LSTM and SVR models to evaluate their effectiveness [11].

Recently, some researchers have BiLSTM, which is short for bi-directional Long Short-Term Memory, and its layers integrated to Transformer block to construct a joint modeling framework and lead to improvements in accuracy [12-15]. So the author uses BiLSTM-Transformer model to get the forecasted returns of alternative assets, and the predicted values are input into the MV model to construct the portfolio. Transformer model and LSTM model are also used as comparison. In the 270-period empirical evidence with six constituent stocks of US30 index, the BiLSTM-Transformer model outperforms the LSTM and Transformer models in terms of forecasting ability, which generated MV model investment performance significantly.

2. Data

The data in this article is derived from Wind database. 6 of US 30 constituent stocks are selected for closing prices, including AAPL.O, BA.N, CSCO.O, DIS.N, HON.O and IBM.N (See Table 1). And the term of a single investment is 1 day. Within the period of February 7, 2022 to March 6, 2023, 270 phases were invested in total. The period number M of observation historical data in each sliding window is 100 days, and the input sequence length m of deep learning model is 5 days.

Table 1: Descriptive statistics of stock closing price data.

Name/Statistics	Mean	Std	Skewness	Kurtosis
AAPL.O	178.3769	153.9492	1.3568	4.1531
BA.N	125.7097	93.7515	1.3581	3.9629
CSCO.O	28.3891	12.122	0.9416	2.7494
DIS.N	67.0637	46.0867	0.7405	2.4431
HON.O	88.0953	58.9945	0.8375	2.4333
IBM.N	130.5815	36.3773	0.2734	2.1729

From Table 1, it can be seen dispersion degree and mean value and kurtosis of the data in ‘AAPL.O’ are the largest, and those in ‘CSCO.O’ are the smallest. The share price data in IBM.N are the most evenly distributed, and those in AAPL.O and BA.N are the most unevenly distributed, which is concentrated on the right side of the axis of symmetry in their distributed graph.

3. Methods

3.1. Transformer-Based Prediction Model

Ashish et al. developed a Transformer model that utilizes a multi-headed self-attentive mechanism [16]. In contrast to conventional sequence models like recurrent neural networks, the attention mechanism incorporated in the Transformer enables it to capture distant contextual information across the sequence, thereby enhancing accuracy and efficiency.

Before the emergence of transformer models, BiLSTM acted as a usual architecture to translate neural machine and answer questions. In order to enhance the performance of these models, Huang et al. [12] have explored the combination of BiLSTM and transformer techniques to create a more powerful architecture. Likewise, Wu et al. [15] have developed a similar BiLSTM-Transformer architecture to make stock closing price prediction precise and accurate.

The BiLSTM-Transformer model structure for time series forecasting is shown in Fig. 1. To enhance the extraction of time series information from stock data, it uses Bi-LSTM recurrent neural network for position encoding replacement in the original Transformer model. This allows the Transformer to capture the time series relationships between input stock data by extracting time series features from the entire data [15]. The encoder maps the input sequence into a latent representation, and the decoder is replaced with a full-connection layer for final data prediction (See Fig. 1).

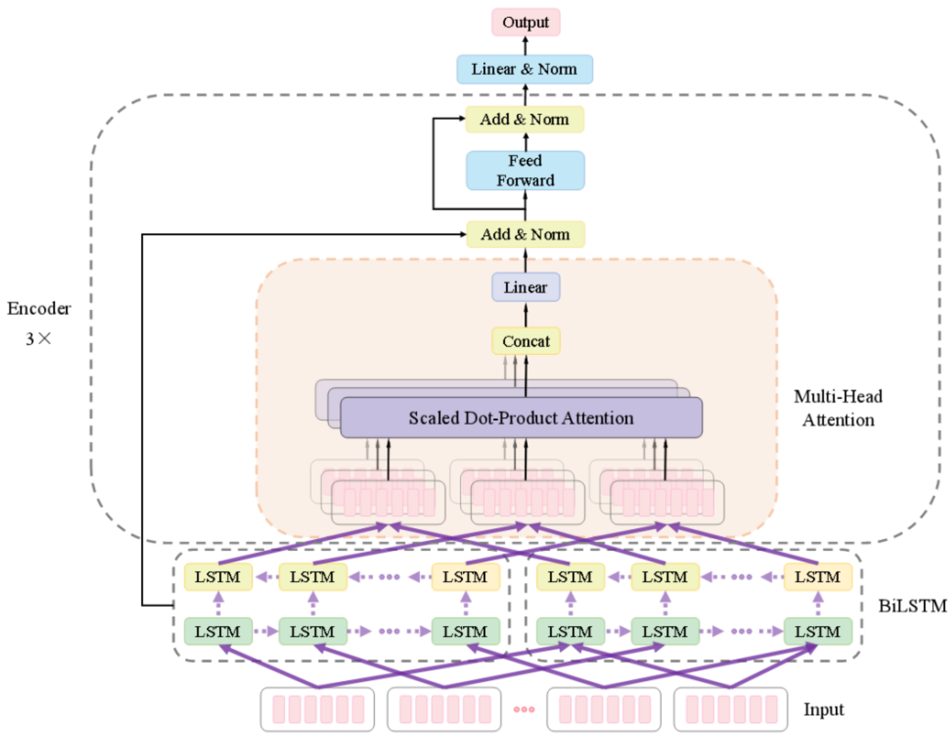

Figure 1: Details of the BiLSTM-Transformer Model.

3.2. Maximum Sharpe Ratio Model

The maximum Sharp ratio model is a special case of Markowitz Mean-Variance model, which acts as optimal variance at minimum level, controls risk by reducing the volatility of the yield, but cannot bring significant expected returns. The maximum Sharpe ratio model in Equation (1) is an optimization approach that comprehensively considers risk (variance) and return, that is, seeks to minimize risk and maximize return. Among them: E(Rp) refers to the expected annualized return on investment

DOI: 10.54254/2754-1169/26/20230573

portfolio, which is the expected excess return vector of alternative assets given by each forecast model; Rf refers to the annualized risk-free interest rate; σp refers to the standard deviation of portfolio annualized return, using the historical excess rate of return of each asset as an estimate;ω refers to the weights vector allocated by investors to investments.

$$\max \frac{E(R_p) - R_f}{\sigma_p} \tag{1}$$

$$s.t.\, \omega^T * 1_n = 1 \tag{2}$$

$$\omega_i \geq 0\,, \forall i \in n \tag{3}$$

4. Results

4.1. Closing Price Prediction

To varify BiLSTM-Transformer model efficiency for predicting stock closing prices, this study utilizes 5578 data points from 2001/01/02 to 2023/03/06 as the original sample for prediction. The training and test sets are divided in a 9:1 ratio. Additionally, a sliding window with a time step of 5 is constructed, meaning that the previous 5 days' stock data is utilized as input for time-series rolling prediction. The pre-processed data is fed into the model, where temporal features are first extracted using a Bi-LSTM recurrent neural network, followed by feature extraction using a 3-layer encoder. Finally, a fully-connected layer is added to transform the data dimension to 1-dimensional for the output of stock closing price prediction. Parameter set for the model are as follows in Table 2.

Table 2: Model parameters.

Parameter name	Parameter	Parameter name	Parameter
Number of BiLSTM neurons	64	Loss function	MSE
Number of multi-heads h	4	Activation functions	sigmoid
Weighting Matrix Dimensions d_k	64	Optimizers	Adam
Number of encoders	3	Number of training rounds	50
Encoder input size d_model	300	batch	712
dropout	0.2	Time step	5
Number of fully connected layer nodes	128	Training set: Test set	9：1

4.2. Comparative Analysis

For the evaluation of model performances in the stock closing price prediction problem, LSTM and Transformer model are selected for cross-sectional comparison.

Four commonly used error evaluation metrics are used to evaluate models: MAPC, which is short for Mean Absolute Percentage Error and MSE, which is short for Mean Square Error, MAE, which is short for Mean Absolute Error, and Determination Coefficient (R^2).

If the MAPE, RMSE, and MAE values are smaller, it indicates that the prediction error is smaller and the prediction is better. Besides, the large value of R^2 leads to a better prediction performance of the model.

DOI: 10.54254/2754-1169/26/20230573

Table 3 shows the models' average forecast error across six stock test sets. As can be seen:

(1) From the overall perspective of the four indicators, the average prediction error of BiLSTM-Transformer model is the smallest, and the accuracy is the highest.

(2) From the perspective of MAPE, MSE and MAE, Transformer is inferior to other models; however, as for R^2, Transformer is superior to LSTM and close to BiLSTM-Transformer (See Table 3).

Table 3: Comparison of closing price forecast results of alternative models.

	MAPE/%	MSE/ $	MAE/ $	R^2
LSTM	2.57	4.4285	3.4632	0.8774
Transformer	2.60	4.8296	3.8409	0.8961
BiLSTM-Transformer	2.33	3.9175	3.1328	0.8992

4.3. Performance of Portfolio

Through the above prediction models, we can obtain the predicted closing price for each day of a 270-day investment period, and then obtain the corresponding predicted rate of return. In this paper, the variance-covariance matrix of the returns of each alternative stock in the past 100 days is used as the estimated value of Σ of the MV model, and the expected returns of alternative stock derived from the three prediction models of BiLSTM-Transformer, Transformer and LSTM are used as the input of the MV model to construct the portfolio for every term. As can be seen from Fig. 2 and Table 4, BiLSTM-Transformer model has advantages in improving the investment performance of MV model.

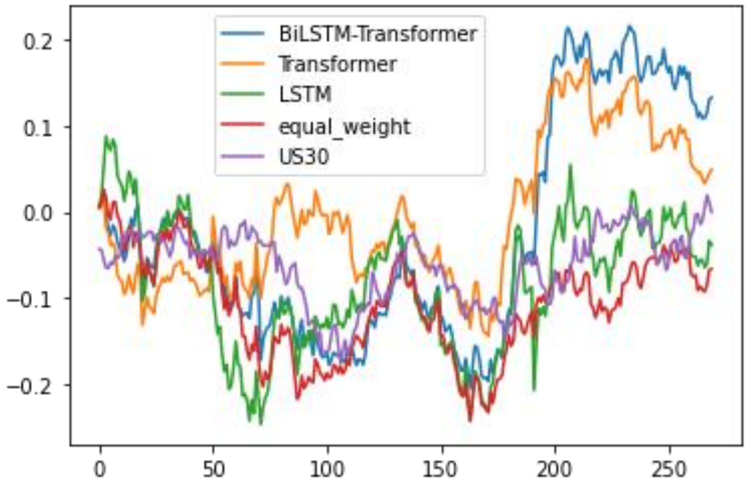

Figure 2: Cumulative returns of portfolio constructed by MV model.

Table 4: Investment performance of portfolio constructed by MV model.

	Cumulative return (%)
BiLSTM-Transformer	13.29
Transformer	4.91
LSTM	-3.84
Equal weight	-6.65
US30	-0.01

DOI: 10.54254/2754-1169/26/20230573

5. Conclusion

The prediction error of the expected excess return of alternative assets will lead to the deterioration of the investment performance of the portfolio model in the traditional portfolio model. Some studies use machine learning methods to predict the return rate of assets, thus improving portfolio model performances. The author has adopted a model with transformer which was improved by Bi-LSTM to predict expectation of alternative asset in return rate, and the predicted value is input into the MV model to build a portfolio. The results are compared with LSTM and Transformer model. In the 270 phase investment empirical study, which takes 6 stocks of US30 index as alternative assets, Transformer model shows better forecasting ability than LSTM and Transformer model, and significantly improves the investment performance of MV model, achieving highest cumulative gain finally. The paper gets to conclusions which expands theoretical research of investment portfolio. The combination of BiLSTM-Transformer model and MV model is also applicable to the investment practice in the real market environment.

This study has certain limitations as it only utilizes basic historical returns as input features. Former studies have found meaning of technical indicators. Thus, indicating future studies to highlight effective input feature identification for model prediction training and MV model performance improvement on daily trade investment. Additionally, the higher the turnover, the more challenges the model will face. This study has not taken transaction costs into consideration. Therefore, future work should also explore the impact of transaction costs on the model's performance.

References

[1] *Markowitz, H.: Portfolio selection. The Journal of Finance 7(1), 77-91 (1952).*

[2] *Dai, Y.L.: Analysis and evaluation of Markowitz model. Financial Research (9) (1991).*

[3] *Freitas, F. D., Souza, A. F. D., Almeida, A. R. D.: Prediction-based portfolio optimization model using neural networks. Neurocomputing 72, 2155–2170 (2009).*

[4] *Hao, C. Y., Wang, J. Q., Xu, W.: Prediction-based portfolio selection model using support vector machines. In Proceedings of sixth international conference on business intelligence and financial engineering, 567–571 (2013).*

[5] *Zhu, M.: Return distribution predictability and its implications for portfolio selection. International Review of Economics & Finance 27, 209–223 (2013).*

[6] *Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., et al.: Decision-making for Financial Trading: A Fusion Approach of Machine Learning and Portfolio Selection. Expert Systems with Applications (115) (2019).*

[7] *Andriosopoulos, D.: Computational Approaches and Data Analytics in Financial Services: A Literature Review. Journal of the Operational Research Society 70(10) (2019).*

[8] *Baek, Y., Kim, H. Y.: ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. Expert Systems with Applications 113, 457-480(2018).*

[9] *Ozbayoglu, A. M., Gudelek, M. U., Sezer, O. B.: Deep learning for financial applications: A survey. Applied Soft Computing 93, 106384 (2020).*

[10] *Ma, Y., Han, R., Wang, W.: Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications 165, 113973 (2021).*

[11] *Zhang, N., Yan, S., Fan, D.: Return prediction and portfolio optimization based on deep learning. Statistical Research 36(3), 67-79 (2019).*

[12] *Huang, Z., Xu, P., Liang, D., et al.: TRANS-BLSTM: Transformer with bidirectional LSTM for language understanding. arXiv preprint arXiv:2003.07000, (2020).*

[13] *Zhao, Z., Chen, Y., Liu, J., et al.: Evaluation of Operating State for Smart Electricity Meters Based on Transformer–Encoder–BiLSTM. IEEE Transactions on Industrial Informatics 19(3), 2409-2420 (2022).*

[14] *Varun, Y., Sharma, A., Gupta, V.: Trans-kblstm: An external knowledge enhanced transformer bilstm model for tabular reasoning. Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, 62-78 (2022).*

[15] *Wu, X.H., Sun, C., Hao, X.Y.: Stock Closing Price Interval Prediction Based on CEEMDAN-WTD-Bilstm-Transformer Model. Available at Research gate https://www.researchgate.net/, last accessed 2023/4/2.*

[16] *Vaswani, Ashish, et al.: Attention is all you need. Advances in neural information processing systems 30, 5998-6008(2017).*