Optimizing Stock Market Predictions with Hybrid Deep Learning Architectures

Catherine Zhang (yz5609) Tianyu Wu (tw1802) Xintong Xie (xx964) Zhiyu Guo (zg915)

DS-GA 1003 Machine Learning Final Project

INDTRODUCTION

In the past years, modern machine learning structures have been widely applied to the study of financial data, with one major part of it being stock market data. There is a growing trend of using Recurrent Neural Networks (RNN) to study the time series data of the stock market. Within the RNN category, Long Short-Term Memory (LSTM) models are particularly esteemed due to their ability to process extended time series while addressing issues like gradient vanishing and gradient exploding during training.

In the context of LSTM, new approaches involve augmenting LSTM architectures with other DNN structures to enhance their predictive power. Sunny et al. explored the use of Bi-Directional LSTM (BI-LSTM) for predicting stock closing prices, which incorporates learning from reverse data sequences and reportedly achieves lower RMSE with proper hyperparameter adjustments. Meanwhile Lu et al. demonstrated that adding a CNN layer to extract features before LSTM processing can yield better model performance compared to using CNNs or LSTMs alone.

RESEARCH QUESTIONS

Building on previous studies, this research introduces a novel CNN-BI-LSTM model to predict stock prices. This study aims to replicate previous model results, develop and test the new DNN combination, employ systematic model comparison techniques, and analyze learning patterns and inductive biases among different neural network architectures. To be specific, this study focuses on answering these research questions:

1. Do CNN-LSTM and BI-LSTM outperform the traditional LSTM model?
2. Can a more complex DNN architecture, such as CNN-BI-LSTM, provide superior performance compared to its predecessors?
3. What specific features are captured by the CNN and Bi-Directional layers?
4. Why is the integration of these complex DNN structures beneficial?

DATA

Date	Opening Price	Highest Price	Lowest Price	Closing Price	Volume (share)	Turnover (RMB)	Up and Downs	Change (%)
2020-08-28	3346.29	3405.88	3339.65	3403.81	2.71e10	3.79e11	53.70	1.60
2020-08-31	3416.55	3442.72	3395.47	3395.68	3.23e10	4.36e11	-8.13	-0.24

Daily trading data is acquired for the Shanghai Composite Index (000001) covering 7,127 trading days from July 1, 1991, to August 31, 2020. The dataset is split into a training set of the first 6,627 days and a testing set of the latter 500 days. This dataset includes 8 variables for each trading day: opening price, high price, low price, closing price, volume, turnover, daily gains/losses, and percentage change. The input data is a (10, 8) matrix containing data for ten consecutive days covering all eight variables. The target variable is the closing price for the eleventh day.

MODELS

CNN-LSTM

This model features a combination of CNN and LSTM components. This includes an input layer, a one-dimensional convolution layer, a pooling layer, an LSTM hidden layer, and a fully connected layer.

BI-LSTM

This network enhances the traditional LSTM by incorporating both past and future data inputs through a dual-layer architecture consisting of a forward LSTM and a backward LSTM. This configuration allows the BI-LSTM to process sequence information in both directions, enhancing the robustness and accuracy of the model.

CNN-BI-LSTM

The CNN-BI-LSTM prediction model operates similarly to the CNN-LSTM, with the primary difference being that subsequent to the CNN layers, the feature outputs are not channeled into a standard LSTM but rather into a BI-LSTM.

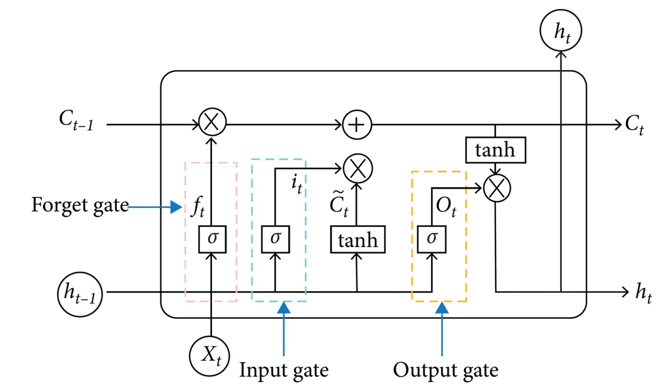


Figure 1. Repeating Module of LSTM

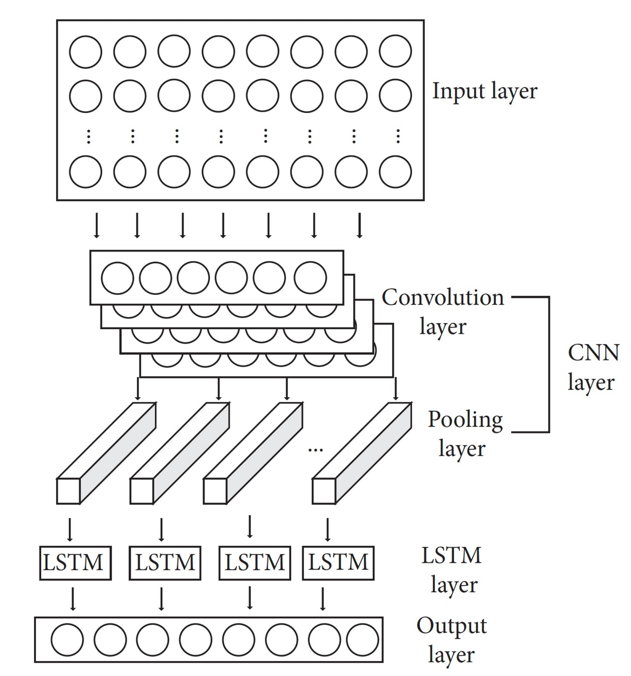


Figure 2. CNN-LSTM Structure

RESULTS

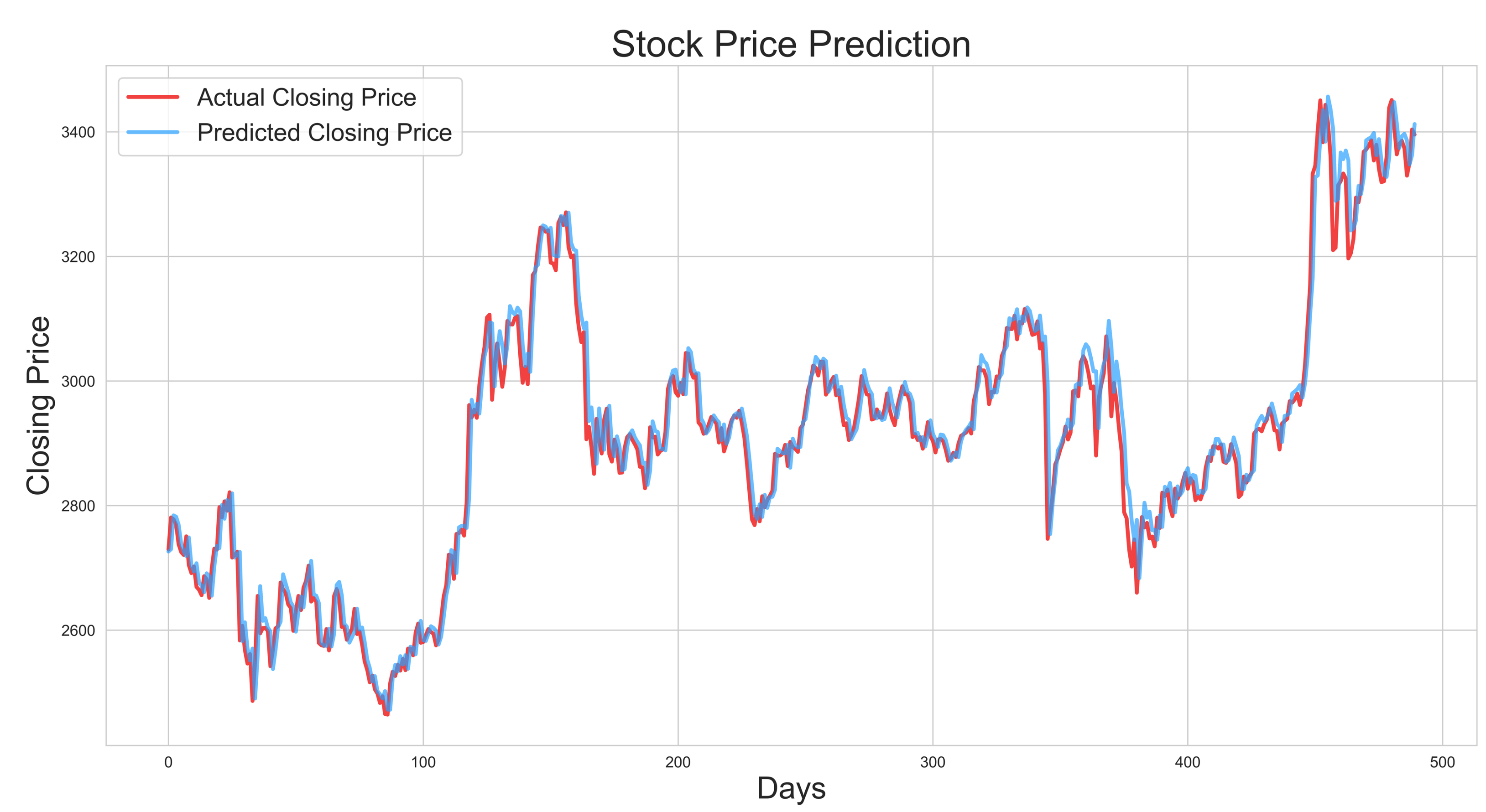


Figure 3. BI-LSTM Prediction on Test Set Closing Price

To replicate results, the exact dataset, model structure and hyperparameter settings are used for training and testing. Figure 3 shows BI-LSTM (one of the best model) prediction on the closing price, using past ten days data as input features.

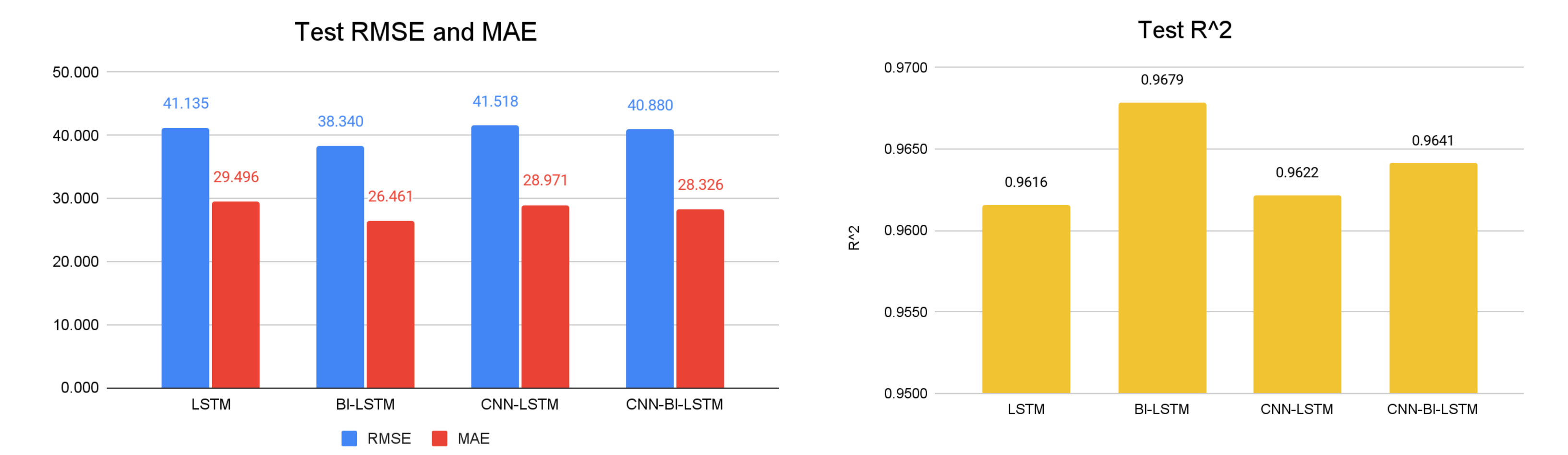


Figure 4. Evaluation Matrices for Model Performance

The result statistics show that the test errors of BI-LSTM and CNN-LSTM are lower, and the R-Squared of are higher than that of LSTM, indicating that BI-LSTM performs better than CNN-LSTM better than LSTM.

As for the novel CNN-BI-LSTM, it outperforms all other models after careful structure and proper hyperparameter adjustments.

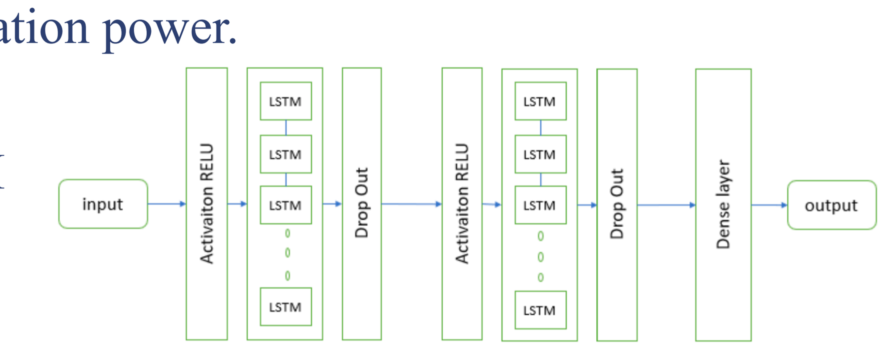
FLAWS IN PREVIOUS PAPER

Overtraining

In both paper, optimal models are acquired with at least 100 training epochs. However, during the replication process, models converge at around 20 epochs.

It indicates previous models suffer from overtraining, which causes overfitting and unnecessary waste of computation power.

Mis-structured Model



The original LSTM and BI-LSTM models include a ReLU activation layer at beginning. When dealing with closing price data after min-max scaling, the ReLU activation is useless.

Mis-understanded Model

Lu et al. employed a 1D convolutional layer with kernel size 1, which suggests that features are extracted only from one day's data. Therefore, this paper to question the effectiveness and theoretical justification of using the CNN layer to "extract the time features of data."

Robustness of Model Performance

Lu et al. show that the CNN-LSTM model outperforms CNN and LSTM. However, with replication, no significant performance difference is found. Therefore, this paper questions whether the CNN-LSTM they proposed truly outperforms other models, or if it is merely a lucky shot.

Reference

Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. *Front. Bus. Res. China, 14*, 13.

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-Based Model to Forecast Stock Prices. *Artificial Intelligence for Smart System Simulation, Volume 2020.*

Sunny, M. A. I., Maswood, M. M. S., & Alharbi, A. G. (2020). Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Giza, Egypt (pp. 87-92).