Creative Economy
2022, VOL. 6, NO. 3, 31-37
DOI: 10.47297/wspceWSP2516-251903.20220603

Research on Stock Price Forecasting Model based on BiLSTM-SVR-based Hybrid Model

Songlin Zhang

Department of Mathematics and Statistics Chongqing Technology and Business University, Chongqing 400067, P. R.China

ABSTRACT

In this study, we propose a combined deep learning model based on BiLSTM-SVR to predict the stock price of TSMC, an upstream company in the semiconductor industry, using three deep learning algorithms, namely Long Term Short Term Memory (LSTM), Support Vector Regression (SVR) and BiLSTM-SVR. For BiLSTM-SVR, we chose the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the goodness of fit (R2) to measure the model performance. The BiLSTM-SVR model combination is completed by using get layer to present the output of last layer as the input of the next layer in the BiLSTM, which is fed into the SVR model for regression fitting. By comparing with the normal LSTM model, we find that the BilSTM-SVR stock price prediction model performs better among all the models in this study.

KEYWORDS

BiLSTM; SVR; Deep learning; Stock price forecasting

1 Introduction

Taiwan Semiconductor Manufacturing Company (TSM) is the world's largest wafer foundry. Its 69.9% market share in the 2022Q1 quarter is pivotal to the trend of the global semiconductor industry[1]. As the impact of supply chain disruptions continues to deepen, the impact of the coronavirus disease epidemic in late 2019 has led to logistical disruptions and shutdowns across the globe affecting the ability to manufacture output of various key chips. The impact of the spread of the epidemic on upstream companies such as TSMC, Samsung, and Intel is often cited as one of the reasons for this chip shortage. At the same time during the epidemic, the sharp rise in crypto-digital currencies worldwide attracted many investors to join the wave of speculation, and a large number of graphics display cards were used for mining. These two main reasons have led to the current chip shortage crisis.

Therefore, analyzing TSMC's stock price trend will help assist investors in their decision-making. For stock price data, the most common is time series. The methods commonly used in this area are the ARIMA model, LSTM model, MLR multiple linear regression, etc.

2 Literature review

In view of the serious hazards posed by the global chip shortage, it is necessary for domestic and international scholars to investigate these issues.

Aishwarya Premlal Kogekar[2] et al. proposed an efficient deep hybrid model, CNNBiLSTM-SVR, for predicting two important water pollutants, and a comparative analysis of the proposed model with four different ones showed that the combination of the CNN-BiLSTM model of the SVR model has better accuracy.

Jiacheng Yang[3] et al. used three machine learning techniques, namely ANN, and SVR, to develop a model for predicting gold price. The results showed that SVR outperformed both models and was the better model for predicting gold in the dataset.

Chulan Ren[4] et al. fully exploited the time series characteristics of the stock data. The work shows that the CBAM-BiLSTM neural network model has higher prediction accuracy for stocks than the ARIMA model, the traditional LSTM model, and the standard BilSTM model, which helps investors to maximize their investment interests.

By listing the RMAE, MAPE, and RMSE of the five models, we can better reflect the advantages and disadvantages of each model by Yiqiu Fang[5] et al. The RMAE values are similar, the MAPE values are 0.04 percentage points lower, the RMSE values are 1% lower, and overall performance is slightly better than LSTM-CNN. The overall performance is slightly better than that of LSTM-CNN.

3 Dataset

The database of Tushare platform is selected for this study and its API is docked to obtain the data. The time range of this dataset is from January 1, 2017, to July 21, 2022, covering 1393 daily closing price observations. An overview of the data is shown in the figure below.

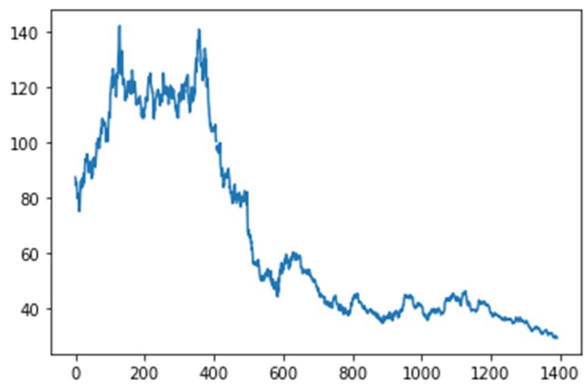

Figure 1 Stock price trend

In this study, the independent variable x and the dependent y will be split into a set oftraining and test set in the ratio of 80% and 20%. And the input and output values will be normalized to avoid bias in the model during training.

In this study, two types of machine learning algorithms, BiLSTM-SVR and LSTM, will be used to classify and predict historical stock prices, respectively.

4 Model implementation

4.1 LSTM

Long-short-term memory is an enhanced temporal recurrent network based on the RNN model[6]. Its basic unit can be divided into one forget gate, an input gate and an output gate.

The forget gate is used to control saving of which should be saved in cell state C^{t-1} time previous to current cell state C^{t}. The formula is defined as follows.

$$\begin{cases} Z_f = \sigma(\omega_{fx}X^t + \omega_{fh}h^{t-1} + b_f) \\ Z_f \odot C^{t-1} \end{cases} \tag{1}$$

The output gate controls how much information in the current cell state C^t is saved to the current output. The predicted value y'at the current moment and another cell state value h are generated after the output gate. Z_o in the output gate is defined as follows.

$$Z_o = \sigma(\omega_{ox}X^t + \omega_{oh}h^{t-1} + b_o) \tag{2}$$

Where h^t is calculated from the cell states C^t and Z_o calculated after input and defined as follows.

$$h^t = \phi(C^t) \odot Z_o \tag{3}$$

The current h^t at this time, in addition to transferring to the next moment as the input unit state, also needs to be calculated again as the output of the current moment. The formula is defined as follows.

$$\hat{y}^t = \omega_y h^t + b_y \tag{4}$$

4.2 BiLSTM

For example, LSTM can only deliver temporal features in the order of past to future for temporal data sets. The bi-directional Long Short-Term Memory (BiLSTM) can solve this problem.

BiLSTM is a reverse LSTM on top of an LSTM, which consists of forward and reverses LSTMs, taking into account the influence of correlation between data on temporal features in addition to forwarding temporal order, and the forward and reverse computations are linearly fused and output again, which can make the final results of the model more accurate[7]. BiLSTM network is as follows shown below.

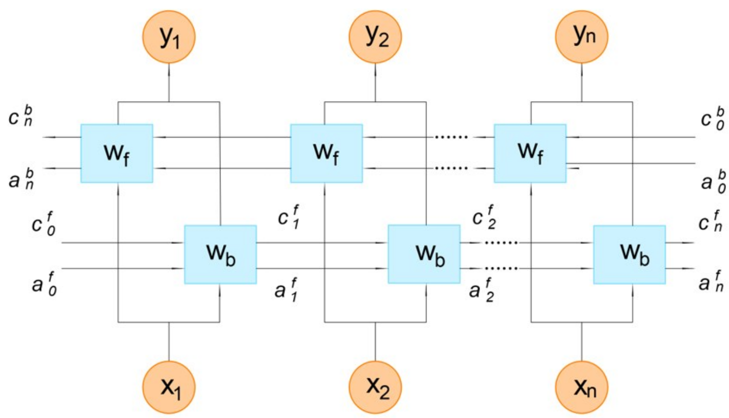

Figure 2 The structure of BiLSTM

4.3 SVR

Support Vector Regression (SVR), which is an important application branch of SupportVector Machine (SVM) in the binary classification problem[8], SVM is the classical machine learning algorithm, which is centered on finding the appropriate classification plane and finding the individual classification sample points that are closest to that hyperplane.

Since the data set is linearly divisible, then there exists a hyperplane defined as.

$$\omega^T x + b = 0 \tag{5}$$

The distance from each sample $x^{(n)}$ in the data set to the hyperplane is defined as follows.

$$\gamma^{(n)} = \frac{\left|\omega^T x^{(n)} + b\right|}{\|\omega\|} = \frac{y^n\left(\omega^T x^{(n)} + b\right)}{\|\omega\|} \tag{6}$$

The objective of the SVM is to find a hyperplane (ω^*, b^*),such that γ maximizes which is defined as.

$$\max_{\omega,b} \rightarrow \gamma \tag{7}$$

$$s.t. \frac{y^n\left(\omega^T x^{(n)} + b\right)}{\|\omega\|} \geq \gamma, \forall n \in \{1,\cdots,N\} \tag{8}$$

The distance from sample $x^{(n)}$ to the split hyperplane does not change because we deflate both $\omega \rightarrow k\omega$ and $b \rightarrow kb$, so that the following conditions are satisfied.

$$\|\omega\| \cdot \gamma = 1 \tag{9}$$

$$s.t. y^{(n)}\left(\omega^T x^{(n)} + b\right) \geq \gamma, \forall n \in \{1,\cdots,N\} \tag{10}$$

SVR enables more flexibility in defining the maximum acceptable error rate in the model, which is the parameter ε.The function of SVR is defined as follows.

$$f(x) = w^T \phi(x) - b \tag{11}$$

Where $f(x)$ represents the decision function. The objective of SVR is to find a function $f(x)$, when the variation of $f(x)$ is maximum at all target values of ε, and to make it as flat as possible., w represents the weight vector, ϕ represents the nonlinear transformation in N-dimensional space, x represents the feature vector, and b represents the bias term. The objective of this function is to find suitable values for w and b that satisfy the evaluation of the input variable x with minimum regression risk which is defined as follows.

$$Reg(f) = C\sum_{i=0}^{l} \tau\left[f(x_i) - y_i\right] + \frac{1}{2}\|w\|^2 \tag{12}$$

Where C represents the constant and τ represents the cost function to verify the flatness of $f(x)$ and the difference between the target and maximized values. The weight vector w can be defined as follows.

$$w = \sum_{i=0}^{l}(a_i - a_i^*)\phi(x_i) \tag{13}$$

Substitute the value of w into the mathematical function of SVR. The defining equation transforms as follows.

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)k(x_i, x) + b \tag{14}$$

Where $k(x_i, x)$ represents the kernel function of the model, and here ReLU is chosen as the kernel function.

4.4 BiLSTM-SVR

In BiLSTM the last layer output results are presented as the input to the next layer, which is fed into the SVR model for regression fitting. A multi-output wrapper is used, with many-to-many predictions, to predict 3 future time points using 14 past time points. The schematic diagram of BiLSTM-SVR is shown below.

4.5 Training process

In this study, a sliding window forecast is used, where the window length is 14 days, the sliding window length is 1 day, and the forecast period is set to 3 days. The initial window value of time (days) is set to $x_1 \in (-1,14)$ and the initial window value of stock price (USD) is set to $y_1 \in (-1,3)$. The comparison of the input window and the forecast length is shown in the following figure.

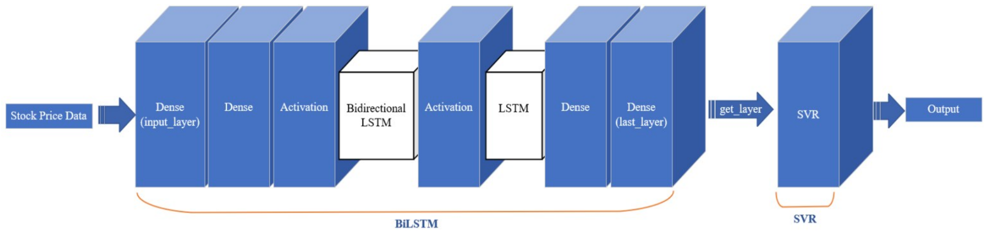

Figure 3 The structure of BiLSTM-SVR

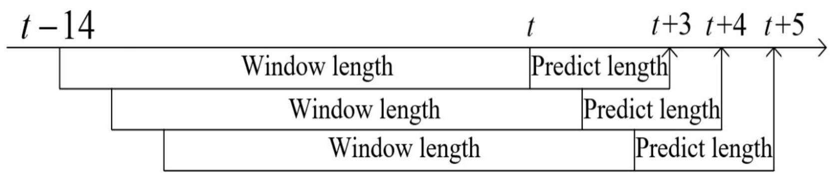

Figure 4 Sliding forecast window

The LSTM part of the model contains 3 Dense layers, 1 Activation layer, and 1 Bidirectional layer. Since the output data is a regression type problem, ReLU is chosen for the activation function layer.

To adapt the data to LSTM, both training and testing datasets are reconstructed from data sequences into a supervised learning framework, the input datasets are reshaped into a two-dimensional format, and the model parameters are selected as shown in the following table.

Table 1 Parameters specification

MODELS	PARAMETERS
LSTM	units=64, epochs=100, batch_size=45
BiLSTM	units=128, epochs=100, batch_size=45
SVR	Kernel=linear, C=3, epsilon=0.1

4.6 Performance benchmarks

In this study, there are various performance benchmarks to measure the merit of model fit. For a more comprehensive representation in this study, we chose root mean square error (RMSE), mean absolute error (MAE), and goodness-of-fit (R2).

To compare the RMSE measures the deviation between the observed and true values. It is commonly used as a measure of the prediction results of machine learning models, which is defined as follows.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2} \tag{15}$$

The mean absolute error is the average of the absolute errors. It can better reflect the actual situation of the prediction value error which is defined as follows.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}_i - y_i\right| \tag{16}$$

The R2 statistic is the proportion of the variance of the dependent variable that can be explained by the independent variables through the regression relationship. The value range is (0,1), and the more R2 is close to 1, the greater the proportion of the sum of squares of the regression to the total sum of squares, which means the better the fit of the regression. R2 is defined as follows.

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\overline{y}_i - y_i)^2} \tag{17}$$

5 Results

After the model is trained, its fit is checked using a test set. This part of the data is not included in the training of the model and is independent of the training results so that it can be measured by a performance benchmark. The results obtained are shown in the following table.

Table 2 Parameters selection

MODELS	BiLSTM-SVR for close data	LSTM for close data
RMSE	0.121	0.128
MAE	0.085	0.087
R2	0.988	0.986

From the table, we can see that the RMSE value of the normal LSTM is 0.128, which is about 5.16% higher than that of, and the MAE value of the normal LSTM is 0.087, which is about 2.54% higher than that of BiLSTM-SVR, which is a significant improvement in accuracy. However, the combined BiLSTM-SVR model in the training and test sets shows a further improvement in accuracy compared with the normal LSTM model.

The MAE of the BiLSTM-SVR is higher than the measured value of the ordinary LSTM, which we believe is related to some unfavorable outliers that lead to slightly lower performance in this index than the ordinary LSTM. Overall, the BilSTM-SVR stock price prediction model performs better than all the models in this study.

After the model starts training. Early on, the learning rate is very high, leading to a significant decrease in step size. As the number of points decreases, the learning rate also decreases, leading to a decrease in the decreasing step size. Although the combined BiLSTM-SVR model has a slightly lower MAE index on the training set than the general LSTM model, it is found that there is room for improvement on the test set. In particular, very low or very high yield sequences can be effectively eliminated when the correct threshold is established. Ultimately, the comparative results of the model fits are shown below.

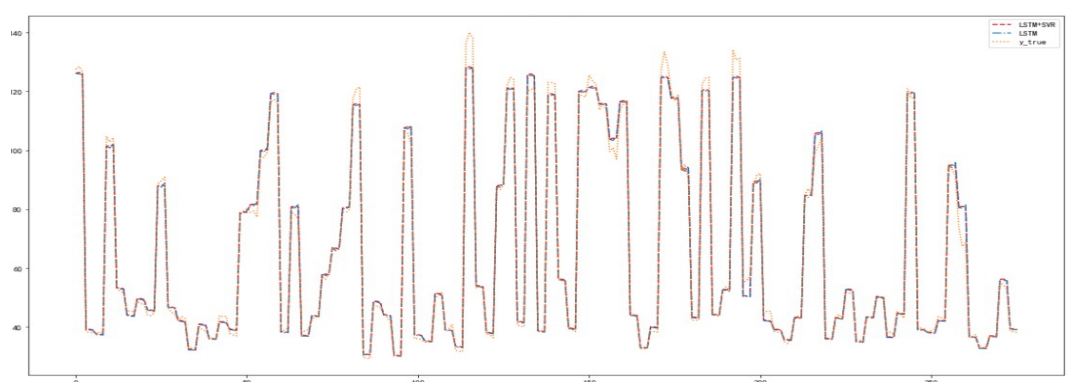

Figure 5 The comparison of BiLSTM-SVR, LSTM and original data

6 Conclusion

In this study, we combined the BiISTM-SVR model in a pioneering way and applied it to stock price prediction for the first time.

A combined BiLSTM-SVR model and a normal LSTM model are built. During the training, BiLSTM-SVR performs better in RMSE and R2 and is slightly inferior to LSTM in MAE performance.

The combined BiLSTM-SVR model performs very well in the training process, with an accuracy of about 5.16% higher than that of the normal LSTM, because the soft spacing band of the SVR model reduces the impact of outliers that are detrimental to the prediction results. Therefore, in the future, more and more models will apply LSTM iterations to adjust the weights, and then use SVR to eliminate the effects of outliers.

About the Author

Songlin Zhang is undergraduate of Chongqing Technology and Business University of Data Science and Big Data Technology, and his research field is data analytics.

References

[1] Longxing Wang. (2019) .Global semiconductor materials market analysis. Integrated Circuit Applications, 1: 1-4.

[2] Kogekar A.P., Nayak R., & Pati U.C. (2021). A CNN-BiLSTM-SVR based deep hybrid model for water quality forecasting of the river Ganga. In: 2021 IEEE 18th India Council International Conference (Indicon). Guwahati, India. pp. 1-6.

[3] Yang J., De Montigny D., & Treleaven P. (2022) . ANN, LSTM, and SVR for gold price forecasting. In: 2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr). pp. 1-7.

[4] Chulan Ren & Caiyan Song. (2022) Research on financial data prediction based on two-way recurrent neural network. Network Security Technology & Application, 4: 53-55.

[5] Yiqiu Fang, Zhuang Lu & Junwei Ge. (2022). Forecasting stock prices with combined RMSE loss LSTM-CNN model. Computer Engineering and Applications, 9: 294-302.

[6] Xin Lin & Xiaodong Zhu. (2022). LSTM stock price prediction model based on attention mechanism. Journal of Chongqing Technology and Business University(Natural Science Edition), 2: 75-82. DOI: 10.16055/j. issn. 1672-058X.2022.0002.011.

[7] Yi Meng & Qingjuan Xu. (2021). Stock forecast based on CNN-BiLSTM and attention mechanism. Journal of Nanning Teachers Education University(Natural Science Edition), 4: 70-77. DOI: 10.16601/j.cnki.issn2096-7330.2021.04.010.

[8] Jinhai Yao. (2022). Study on stock index prediction based on ARIMA and information granular SVR combination operations. Research and Management Science, 5: 214-220.