

Comparison of ARIMA, ANN and LSTM for Stock Price Prediction

Qihang Ma¹

¹Sun Yat-sen University, Guangzhou, Guangdong, P. R. China

Abstract—The prediction of stock prices has always been a hot topic of research. However, the autoregressive integrated moving average (ARIMA) model commonly used and artificial neural networks (ANN) still have their own advantages and disadvantages. The use of long short-term memory (LSTM) networks model for prediction also shows interesting possibilities. This article compares three models specifically through the analysis of the principles of the three models and the prediction results. In the end, it is believed that the LSTM model may have the best predictive ability, but it is greatly affected by the data processing. The ANN model performs better than that of the ARIMA model. The combination of time series and external factors may be a worthy research direction.

1 INTRODUCTION

Predicting the future seems to be an ability that everyone wants to possess, especially when it can bring benefits. This may be why stock price forecasting is so popular. Although proponents of efficient market hypothesis believe that stock price fluctuations are impossible to predict, a considerable number of researchers claim that some models are acceptable as long as they can produce predictions with considerable accuracy. The most commonly used models are artificial neural networks (ANN) and autoregressive integrated moving average (ARIMA) models. Of course, in recent years, some researchers have proposed that long short-term memory (LSTM) networks has higher prediction accuracy. This article will focus on the establishment principles of these three models and their differences in the prediction results of the stock price of the same company in a certain period, hoping to provide some convenience for later researchers.

The ARIMA model which is used for analysis and prediction has been considered as a very effective prediction technique, especially in social sciences. The prediction does not need to assume any underlying models or related equations. Because ARIMA's forecast results are derived from the values of the input variables and error terms. [1] But limited to it is a linear regression model, ARIMA may have some deviations when facing complex nonlinear practical problems. However, in terms of short-term forecasting, the linear models usually outperform the complex structural models. [2]

ANN is a data-driven adaptive model, with almost no prior assumptions. [3] It is used as a predictive model and is widely used in many fields including finance, commerce, and engineering. The prediction of ANN is based on the results obtained from the original data to make broad observations, and then infer the potential part of the whole. Unlike the ARIMA model, it is very

effective in solving nonlinear problems. The changes in the stock market are also non-linear. Therefore, ANN can provide better results in terms of stock price prediction, compared with traditional models. [4]

Although ARIMA and ANN have been widely used in stock price forecasting, these models cannot measure the continuity of evolving price trends. LSTM is a variant of recurrent neural network. Unlike other methods, its feedback connection makes it easier to find development trends through the back propagation of current historical prices and current prices. However, since the LSTM model is rarely used in previous studies, and few research institutions conduct thorough preprocessing of the data, the performance of LSTM cannot be well demonstrated. This is also an important reason why it has not been widely used. [7]

In the following sections, the differences in the process and results of various models will be further discussed.

2 ANALYSIS

2.1 ARIMA Model

2.1.1 Theoretical introduction of ARIMA model.

The ARIMA (p, d, q) model is a time series analysis model proposed by the American statistician Box GE P and the British statistician Jenkins GM in the 1970s. It is also called the Box-Jenkins methodology. It has the effective ability to generate short-term forecasts. In short-term forecasts, it constantly outperformed complex structural models. The ARIMA model for stationary time series is also called the ARMA (p, q) model. In the ARMA model, the future value of a variable is a linear combination of the past value and the past error, expressed as follows:

* Corresponding author: qihang_ma@qq.com

$u_t = \omega_0 + \omega_1 u_{t-1} + \dots + \omega_p u_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ (1)
 where, u_t is the actual value at t , $\{\varepsilon_t\}$ is the white noise sequence, p and q are integers which are called autoregressive and moving average, respectively. When dealing with a non-stationary time series, certain processing is required to generate a new stationary series. For non-stationary time series with short-term trends, difference can be made to make non-stationary series into stationary series. The simplest equation of this type of model is ARIMA (1, 1, 1), which can be expressed as:

$$(1 - B)(1 - \omega_1 B)u_t = (1 - \theta_1 B)\varepsilon_t \quad (2)$$

where, B is the lag factor, like $Bu_t = u_{t-1}$, $B^p u_t = u_{t-p}$.

2.1.2 Tools and methods for time series analysis

2.1.2.1 The order determination method of ARMA model. The methods for determining the values of p and q in equation (1) mainly include the autocorrelation and partial correlation function order determination method and the AIC criterion, and the AIC criterion is the most widely used. Firstly, the model should be chose by examining the properties of the autocorrelation and partial correlation coefficients of stationary time series. If the autocorrelation coefficients are tailed and the partial autocorrelation coefficients show p -order truncation, select the ARMA (p , 0) model. If the autocorrelation coefficient has q -order truncation and the partial autocorrelation coefficient has tailing, the ARMA (0, q) model is selected. In other cases, select the ARMA (p , q) model. Secondly, when it is difficult to determine

the order based on the correlation coefficient, AIC can be used to determine the order of the model. When p and q reach a certain logarithmic value, AIC (p , q) has a minimum value. At this time, p and q are the best model orders.

2.1.2.2 Stationarity test of ARMA model — ADF Test

$$\Delta u_t = c + \delta u_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta u_{t-i} + \varepsilon_t \quad (3)$$

First, perform regression on equation (3) and divide the estimated $\hat{\delta}$ by its standard deviation to obtain a statistic τ . Then check the ADF critical table to see if the hypothesis ($\delta = 0$) can be rejected. If the absolute value of τ exceeds the absolute value of the ADF critical value, the hypothesis can be rejected, that is, the time series is considered to be stationary. On the contrary, you cannot reject the hypothesis that the time series is non-stationary.

2.2 ANN Model

2.2.1 Theoretical introduction of ANN model As a model widely used in approximating functions and predictions, one of the most significant advantages of ANN is that it is a general approximation model, which means that it can approximate and approximate many functions. Its ability comes from the parallel processing of data rather than the pre-set model.

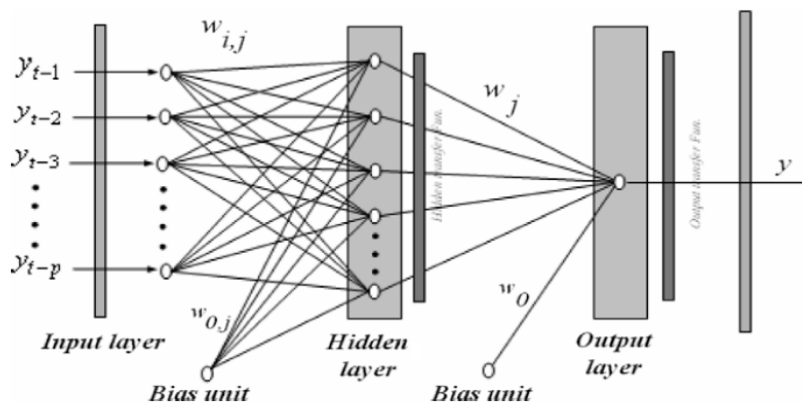


Figure 1. . Neural Network Structure[3]

For models using for time series modeling and forecasting, single hidden layer feedforward network is the first choice. [8] It is characterized by a network composed of three layers of simple processing units connected through acyclic links. The mathematical relationship between the output variable value and the input variable can be expressed as:

$$y_t = w_0 + \sum_{j=1}^q w_j g(w_{0,j} + \sum_{i=1}^p w_{ij} y_{t-i}) + \varepsilon_t \quad (4)$$

where, $w_{ij} (i = 0, 1, 2, \dots, p; j = 0, 1, 2, \dots, q)$ are model parameters called the connection weight, p is the number of input nodes, and q is the number of hidden nodes. y_t

is the output variable, and $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ are the input variables.

Under normal circumstances, the activation function has several forms that can be used, and its type is indicated by the condition of the neurons in the network. Of course, neurons in the input layer generally do not have an activation function, because their role is mainly to transmit the input variables to the hidden layer. In the output layer, the linear function are widely used because they are unlikely to introduce distortion into the output of the predicted value. Logistic function (5) and hyperbolic function (6) are usually used as the transfer function of the hidden layer. Besides, other functions are also possible, such as linear and quadratic functions. The modeling procedure corresponding to each function will be different.

$$L(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (6)$$

Therefore, the ANN model represented by equation (1) is actually a non-linear function between past observations and future values. In other words, the ANN model is actually equivalent to a nonlinear autoregressive model. When the output node q of the ANN model is large enough, it can approximate any function to the hidden nodes, which makes it unexpectedly powerful. Of course, even an ANN model with a simple structure can get a good prediction effect. However, the construction of the ANN model may have an overfitting effect, which will make its generality to out-of-sample data worse.

2.2.2 Selection of the number of input and output nodes. The choice of the value of q is mainly depending on the data, and there is no effective rule that can be referred to in the process of determining this parameter. In addition to selecting an appropriate number of hidden nodes, another important decision to use ANN for time series prediction is to select the lagging observation value p and the dimension of the input vector. Its selection plays a very important role in determining the autocorrelation structure of time series.

There are many different ways to choose p . It can be roughly divided into the following categories. The first is the empirical or statistical method of studying internal parameters and selecting appropriate values according to the performance of the model. [9] The second is hybrid methods such as fuzzy inference. [10] Of course, it is also a feasible way to determine the construction or pruning algorithm of adding or deleting neurons to the original architecture by observing the change value of the system state after deleting or adding neurons. [11] However, although these methods are very complicated, they still cannot guarantee to provide the best solution to all actual

forecasting problems. Testing a large number of networks with different inputs and hidden units, calculating their generalization errors are usually performed to get the suitable model. Once both p and q have been specified, the model can be used for training parameters and the estimation process.

2.3 LSTM Model

2.3.1 Theoretical introduction of LSTM model.

LSTM model is a kind of recurrent neural network (RNN), which mainly solves the problem of gradient disappearance that is easy to occur in traditional RNN, so that it can analyze longer time series data. Compared with RNN, the LSTM model adds three memory modules: input gate, output gate and forget gate. The working principle of LSTM is to process the input information at time t , select useful information with a certain probability, and finally extract useful information through the output gate as the state of the final retention layer, and then participate in the calculation of the next time.

After inputting a new set of variable values x_t at time t and the output state h_{t-1} of the hidden layer at the previous time through the operation of the output gate state o_t , the output result h_t of the hidden layer is obtained. The new cell state C_t and hidden layer output will be passed to the next moment to participate in the calculation. The mathematical relationship of the whole process can be expressed as:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (7)$$

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (8)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (9)$$

$$\vec{C}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (10)$$

$$C_t = C_{t-1} \odot f_t + i_t \odot \vec{C}_t \quad (11)$$

$$h_t = o_t \odot \tanh(C_t) \quad (12)$$

where, W_i, W_f, W_o, W_c are the weights of h_{t-1} . U_i, U_f, U_o, U_c are the weights of x_t . b_i, b_f, b_o, b_c are bias factors. σ is the sigmoid function making sure that the values of i_t, f_t, o_t are between 0 and 1. \tanh is the hyperbolic tangent function.

It can be seen from the above equation that the size of h_t is jointly affected by the current cell state C_t and the

information h_{t-1} contained in the hidden state at the previous moment. And W_c , the main reason for the disappearance of the gradient, has no influence on the calculation of the current cell state. Therefore, by adding a gating structure, the problem of gradient disappearance in the training process is effectively reduced, and the accuracy of model prediction is also improved.

2.3.2 Data Processing for Stock Forecast. When applying the LSTM model to predict stock prices, unlike the ARIMA and ANN models, new technical indicators need to be added to the data. [7] When these indicators are added to the model as input variables, the LSTM model can distinguish between temporary price surges or rapid declines and long-term trend reversals. This will help us distinguish real price trends from market anomalies.

These indicators can be roughly divided into four basic types: trading volume, trend, momentum and volatility. Under normal circumstances, OBV will be used to characterize the trading volume, and use the trading volume to determine the strength of the continuing trend. Momentum indicators, such as RSI, can find out the rate of price change in a given period of time, representing the health of the current upward trend. Trend indicators, such as Fibonacci retracements, MACD, are used to find any developing trend reversals in the market.

Of course, like other models, it is necessary to further process the obtained data by using Z-score

standardization. This will help suppress low data fluctuations and amplify higher values.

2.4 Comparison and Discussion

2.4.1 Comparison of ARIMA and ANN Model. Many researchers have conducted comparisons between ARIMA and ANN models for predicting stock prices. In this part, the forecast of DELL's stock price model around 2010 will be presented for a detailed demonstration. [12]

The model choices are ARIMA (1, 0, 0) and ANN (10, 17, 1). By observing the results, we can find that the prediction of the ANN model tends to predict the numerical value of the stock price. The ARIMA model's prediction is directional, which comes from its model assumptions are linear. Although it can be seen from the chart that the ANN graph is closer to the fluctuation of the real stock price, there is no obvious difference in the accuracy of the prediction results of the ARIMA and ANN models, but the ANN model fits the fluctuations of the test data better. When the statistical test of the results of the two models is performed, it also shows that their performance is almost the same. The P values of the two model tests were 0.439 and 0.604. However, despite this, ANN is relatively more appropriate than ARIMA. This result is also roughly in line with the conclusions of other researchers. [13]

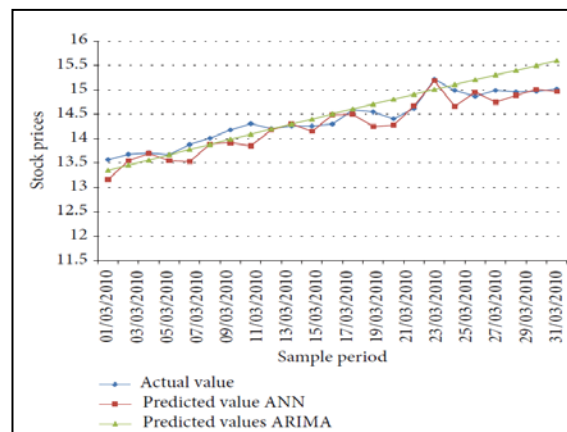


Figure 2. Comparison of ARIMA and ANN model with actual stock price. [12]

Obviously, the superiority of ANN mainly comes from its essentially non-linear model. The price fluctuations of stocks are more suitable for non-linear models than linear models. However, this conclusion may also be overturned when different data are selected. In addition, some researchers have proposed to mix the ARIMA model and GARCH to overcome the shortcomings of linear models while retaining the excellent performance of ARIMA in short-term forecasting. [15]

2.4.2 Comparison of ANN and LSTM Model. There are relatively few studies on stock price prediction using LSTM models. In a few studies comparing the prediction results of LSTM and ANN models, we can conclude that the LSTM model performs better. [7] And this difference may come from the improvement of the LSTM model on the problem of vanishing gradient. It may also come from the addition of indicators to help distinguish between market fluctuations and accidental fluctuations in the input variables. But the ANN model only uses time series to fit the data.

Although the LSTM model shows certain advantages, it has not been studied extensively. This may be because the current data processing is more suitable for the ANN model. The data that is not completely processed is introduced into the LSTM model, which will have a great impact on its performance, making the prediction results unsatisfactory.

3 CONCLUSION

Through the analysis of the establishment process and results of these three models, conclusions can be made. The ANN model is better than that of the ARIMA model, and the performance of the LSTM model may be more due to the ANN. And ARIMA-GARCH can further improve the accuracy of the ARIMA model by improving the white noise sequence.

The disadvantage is that, as we all know, the fluctuation of stock prices is not only related to changes in time, but also related to economic factors, socio-political factors, and the listing of other stocks. Although the LSTM model introduces other variables than the other two models to distinguish market fluctuations and sudden changes, the three models are essentially deduced by using possible relationships in the time series without considering other external factors. This is also the direction in which future research can be further in-depth. Of course, the further development and use of LSTM model in stock price prediction is also a subject of research value.

ACKNOWLEDGEMENT

Professor Coggeshall and Yuhui are very grateful for the inspiration they brought me. Similarly, I am very grateful to my family for their support and encouragement. Olivia also gave good suggestions during the writing process.

REFERENCES

1. G. Tabachnick and L. S. Fidell. (2001). *Using Multivariate Statistics*, Pearson Education, Upper Saddle River, NJ, USA, 4th edition.
2. A. Meyler, G. Kenny, and T. Quinn. (1998). *Forecasting Irish Inflation Using ARIMA Models*. Technical Paper 3/RT/1998, Central Bank of Ireland Research Department.
3. M. Khashei and M. Bijari. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489.
4. Y. Chen, B. Yang, J. Dong, and A. Abraham. (2005). Time-series forecasting using flexible neural tree model. *Information Sciences*, vol. 174, no. 3-4, pp. 219–235.
5. F. Giordano, M. La Rocca, and C. Perna. (2007). Forecasting nonlinear time series with neural network sieve bootstrap. *Computational Statistics and Data Analysis*, vol. 51, no. 8, pp. 3871–3884.
6. A. Jain and A. M. Kumar. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing Journal*, vol. 7, no. 2, pp. 585–592.
7. A. Sethia, P. Raut. (2018). Application of LSTM, GRU and ICA for Stock Price Prediction. *Proceedings of ICTIS 2018*, Volume 2.
8. G. Zhang, B. E. Patuwo and M. Y. Hu. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62.
9. P. G. Benardos and G. C. Vosniakos. (2002). Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. *Robotics and Computer Integrated Manufacturing*, 18, 43–354.
10. J. Leski and E. Czogala. (1999). A new artificial network based fuzzy interference system with moving consequents in if-then rules and selected applications. *Fuzzy Sets and Systems*, 108, 289–297.
11. Jiang, X., & Wah, A. H. K. S. (2003). Constructing and training feed-forward neural networks for pattern classification. *Pattern Recognition*, 36, 853–867.
12. A. A. Adebisi, A. O. Adewumi, and C. K. Ayo. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, Volume 2014.
13. J. T. Yao, C. L. Tan, and H. L. Poh. (1999). Neural networks for technical analysis: a study on KLCI. *International Journal of Theoretical and Applied Finance*, vol. 2, no. 2, pp. 221–241.
14. V. R. Prybutok, J. Yi, and D. Mitchell. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research*, vol. 122, no. 1, pp. 31–40.
15. Tang, Him et al. (2003). Finite Mixture of ARMA-GARCH Model for Stock Price Prediction. *CIEF*, 2003, pp. 1112–1119.

