

Stock Prediction and Analysis based on RNN Neural Network

Lei Lin

Division of General Studies, University of Illinois Urbana-Champaign, Champaign, United States

Abstract. The greater the investment, the greater the risk, and building a stock prediction model with high accuracy is of great theoretical significance and practical application for financial investors. It has become a trend to apply artificial neural network to stock prediction. In this paper, we select the Shanghai Composite Index to predict and analyze the stock price. A three-layer neural network is built and the convergence rate is analyzed, and it is obtained that the fitting effect will be more accurate when the selected data is reasonable and has good properties, and finally the change of the stock in a short period of time is obtained. The neural network is feasible and reasonable for stock price prediction, which in turn helps to improve the profitability of stockholders.

Keywords: Stock analysis, neural network, RNN

1. Introduction

Due to the increasingly vigorous development of the market economy, the system of financial services has become more and more perfect. People are gradually investing in financial market and the level of people's demand for financial information is increasing. The stock market is slowly becoming an integral part of the financial industry in general and the securities industry in particular [1]. Surge and slump are two common phenomena in the stock market, and they can directly affect not only a healthy economic development system but also the vital interests of stockholders, which in turn puts the market economy in a disadvantageous position. Along with the informatization of modern economic society and the modernization of the Internet, technological systems such as computers are becoming more and more complete. Information technology provides a very powerful technical basis and tools to solve the problem of financial information. Reasonable and accurate prediction of stock price trends has become the main research and exploration direction for many scholars.

Researchers at home and abroad have conducted various reasonable studies and theoretical analyses of the stock market, using different research methods for the stock market with different characteristics, producing significant effectiveness. Research methods for stock prices can be broadly divided into time series analysis methods, securities analysis methods, gray prediction methods, and artificial intelligence methods [2]. Zhang Xiaochun used logistic regression and LSTM neural network model to predict stock price data [3]. Huang Chaobin used multiple models to predict stock prices and compared the prediction results[4].

Stock price is a time series form of data, the rise and fall of the stock market is related with a variety of good and bad news, showing the state of people's confidence, but there is a certain correlation between the time sequence, so a special type of recurrent neural network (RNN) can be used to model and learn from this time-series related data.

2. Overview of neural network

2.1 Neural network

A neural network is a typical multilayer neural network in which each neuron between the layers is interconnected, which simply means that each neuron in the lower layer is connected to the neuron in each of the upper layers to achieve a connection weight, but there is no connection weight between the neurons in the same layer. The structure of the neural network is shown in Figure 1 [5].

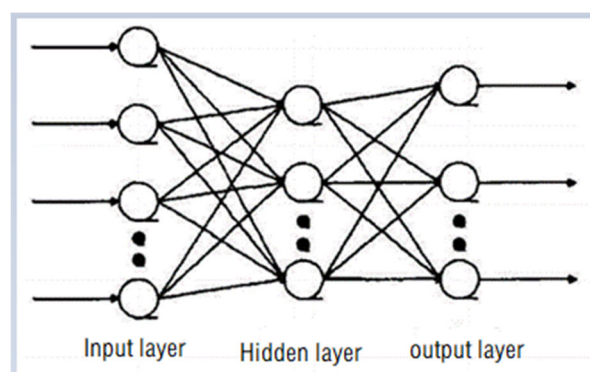


Figure 1. Structure of neural network

2.2 Recurrent Neural Networks

In the traditional neural network model, it is fully connected from the input layer to the hidden layer to the output layer, and the nodes between each layer are connectionless. Therefore, traditional neural network training and testing are slow, and also prone to overfitting, local minimum, and other defects, and this ordinary neural network is incompetent for many problems. Recurrent Neural Networks (RNN) is a deep network model built specifically for sequential data analysis, which has gained wider application in speech recognition, language translation, and picture description [6]. Because of this, RNN have gained great success in natural language processing (NLP). In RNN network systems, different time step lengths have the same one weight and can be cyclically linked across time. Due to the sharing of weights, the temporal parameters in the RNN network system are greatly reduced. In practice, the current state is often assumed to be associated with only a few previous states in order to reduce complexity. The general structure of RNN loops and the looping modules are shown in Figure 2 [7].

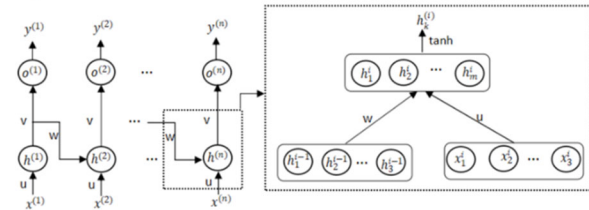


Figure 2. Structure of neural network

3. Stock prediction and analysis

The prediction steps are shown in Figure 3, which mainly include acquiring stock-related data, scaling the data in magnitude, splitting the data into training, and testing sets, constructing RNN models, and model prediction and application.

3.1 Data source

This experiment uses Python to process a stock in the SSE index. The data were selected for the period from October 15, 2018 to October 14, 2022, with a total of 1046 data. The data indicators mainly include: open price, high price, low index, close price of trading data, and volume of constituent securities. Sample of trading data are shown in Table 1.

Table 1. Sample of stock data

Trading Date	OpenP rice	CloseP rice	HighP rice	LowP rice	Volu me	Amount
2018/10/15	38.555	37.348	39.092	37.147	19564292	111620218
2018/10/16	37.281	36.141	37.751	35.94	21409682	117750310
2018/10/17	36.745	36.879	37.214	35.806	24031624	131320280
2018/10/18	36.544	35.739	36.678	35.538	17711684	95611246
2018/10/19	35.001	36.678	36.946	34.666	22024600	119053631

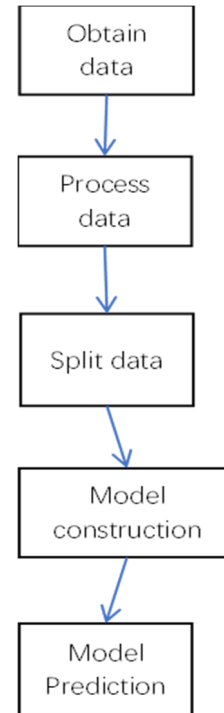


Figure 3. Prediction steps

3.2 Normalization

The main purpose of data normalization is to modify the distribution of the variable data so that the input data can better match the distribution of the output variables [8]. Neural network is a model sensitive to the magnitude of the input data, and large magnitude differences in different fields can affect the training convergence speed and accuracy of the network. In neural network computation, training data are used to learn the neural network and derive the weights, and test data's extrapolation ability to check the neural network; therefore, data are the most essential elements for an effective neural network. The calculation results are affected by the large number of outline indicators due to the large variability of the dimensions among stock indicators, and normalization of the data is required to eliminate this effect. The data is normalized and scaled for better feeding into the neural network and training. In this paper, we use the maximum-minimum normalization method to normalize the data [9]. The linearization of the original data is converted to the range of [0,1] with the following normalization formula, where xmax is the maximum of the sample data and xmin is the minimum of the sample data.

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.3 Split the data

The data set processed in the data preparation phase needs to be partitioned into two parts: a training set and a test set, and the laws obtained in the training phase can be considered reasonable if they can obtain correct results after being tested by the training set [10]. In this paper, we

use the open price of n consecutive days to predict the open price of $n+1$ th day as the target. 80% of the data are selected as the training set and 20% of the data are used as the test set. The loss function is determined along with the data set. The loss function is used to evaluate the degree of difference between the predicted and true values of the model, the better the loss function is, the better the performance of the model is, and the loss function used in different models is generally different. In this paper, the Mean Square Error (MSE) is selected as the loss function and the formula is calculated as follows [11].

$$MSE = \frac{\sum_{i=1}^N [f(x_i) - y_i]^2}{N} \quad (2)$$

3.4 Stock price model construction and forecast

In this paper, we let $n = 5$. Because of the need to make predictions of future values based on historical information. A sliding window data is constructed for the data, and data samples are created along the time series with a time window length of 5, as shown in Figure 4. The RNN model includes input layer, 2 hidden layers, and continuous value output layer. Code is written to implement stock price prediction, and the logic of the code implementation is as follows. As the training process advances, the model loss is continuously optimized, and the optimization and loss decrease rapidly at the initial stage, and then gradually level off. The prediction data obtained by substituting the test set into the model and the original real data are displayed visually, as shown in Figure 5, where the MSE is 0.67 and the fitting effect is reasonable.

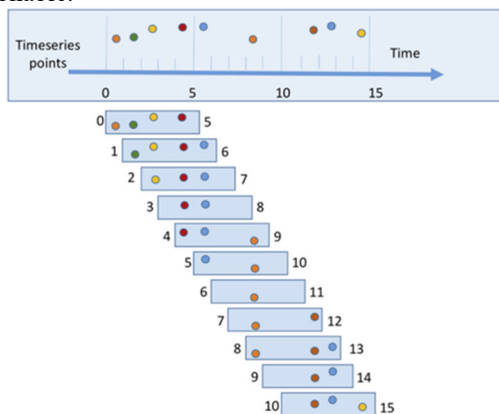


Figure 4. Time window

#Model Training

Preparation of training sets and test sets.
 Put data sets to the input layer.
 The second layer of unfold node is set to 128;
 The third layer of unfold node is set to 64;
 One data output from the fully connected layer.
 The number of epochs is chosen to be 50 considering the machine performance;
 Pack every 20 samples into one batch for training, i.e., the batchsize is 32.
 Start model training.

#Model Test

Test set input model;
 Plotting the results;

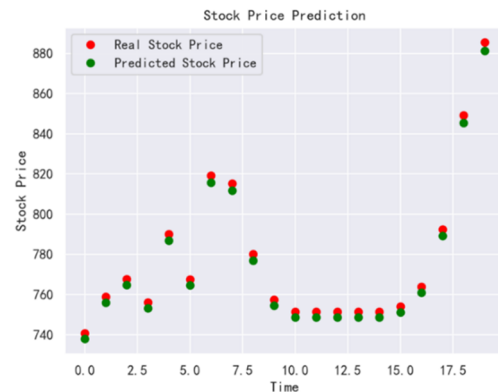


Figure 5. Example of Prediction

The EVS(Explained Variance Score)metricis used to measure the extent to which the neural network model fluctuates on the stock dataset[12].The model is perfect if it takes a value that converges to 1. The smaller the value the less effective it is.Where y is the true value, \hat{y} is the predicted value, and var is the variance.The variance is calculated as 0.99 based on the predicted results and the model results are as expected.

$$\text{explained_variance}(y, \hat{y}) = 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}} \quad (3)$$

Stock prices fluctuate frequently and often quite drastically. The visualization results show that neural network training on stock data is a complex process, and the selection of the number of cycles and nodes affects the training effect, and how to take the parameters needs to be verified by a lot of experiments. The number of nodes in the hidden layer is not as large as possible, and not as small as possible, but a suitable number should be selected based on many experimental results. If the number of hidden layer nodes is too small, the neural network structure will be too simple, and then its learning and classification ability will be reduced. However, too many nodes in the hidden layer will make the neural network structure too complex, overload the network, and reduce the efficiency and generalization ability. Therefore, it is important to train the stock prediction model to ensure the classification ability of the neural network on the one hand, and the generalization ability of the neural network on the other hand. The neural network model designed in this paper can predict the stock data well and make the prediction result very close to the actual value, which provides a meaningful analysis tool for the general investors.

4. Summary

Predicting the movements of stock price has been a hot topic in stock market research. In view of the high correlation and complexity of stock prices, this paper uses

an RNN model to predict a stock price in the SSE index. The feasibility is first explored in terms of both the problem and the method through the knowledge of stocks as well as neural networks. Affected by various factors, the road to study the stock price prediction problem is still long, and there are still many problems to be solved. In future research, we can try to add more indicators, further optimize the structure and parameters of the model, or compare multiple methods in order to explore more accurate prediction. Of course, stock prediction models can only be used as an auxiliary tool for investment decisions, and the actual investment requires manual intervention to make decisions based on different investment strategies. In terms of the long-term use of the same kind of investment model and the changing securities market, model structure and algorithm need to be adjusted according to the actual investment environment and investment strategy changes to get a prediction system that is more suitable for market operation.

References

1. Shi Yanli, Gao Shichen, Wang Juan. Application of artificial neural networks in stock market forecasting. Group Economic Research, 12(2007)1.
2. Chu Wenhua. Application of BP artificial neural network in stock forecasting. Journal of Harbin University of Commerce: natural science edition, 35(2019)4.
3. Zhang Xiaochun, Xu Xiaopeng, Wei Sulin .Application of LSTM neural network in stock price prediction. Computer Knowledge and Technology, 16(2020)39-43.
4. Huang Chaobin, & Cheng Ximing. Research on stock price prediction based on LSTM neural network. Journal of Beijing University of Information Science and Technology (Natural Science Edition), 36(2021)79-83.
5. Fu Shifeng., Zhao Li, Cai Wenjun. Application of BP neural network in stock market prediction. Computer Knowledge and Technology: Academic Edition, 2021.
6. Xi Xiaoya, Qin Hebin, Lu Zhijuan. Research on stock price change prediction based on LSTM neural network model--Baidu stock price as an example. National circulation economy, 16(2022):4.
7. Wang Ding, Zou Yunkai, Tao Yi, et al. Research on password guessing models based on recurrent neural networks and generative adversarial networks. Journal of Computer Science, 44(2021):16.
8. Tang Rongzhi, Duan Huichuan, Sun Haitao. Research on normalization of SVM training data. Journal of Shandong Normal University: natural science edition, 31(2016):6.
9. Ji X, Du S, Wang G. Using Min-Max Normalization to Measure the Differences of Regional Economic Growth—A Case Study of Yulin Area, Shanxi Province. Economy and Management, 2016.
10. Chen, Feng-Xian, Zhao, Jian-Qun. LSTM-based multi-classification prediction of stock prices. Computer Science and Applications, 9(2019):10.
11. Nengroo S H, Lee S, Jin H, et al. Optimal Scheduling of Energy Storage for Power System with Capability of Sensing Short-term Future PV Power Production. 2021.
12. Justin D, Concepcion R S, Calinao H A, et al. Using stacked long short term memory with principal component analysis for short term prediction of solar irradiance based on weather patterns. 2020 IEEE REGION 10 CONFERENCE (TENCON). IEEE, 2020: 946-951.