Research on Prediction and Modeling Method of Financial Time Series based on Deep Learning

Shichen Qiu[1], Fan Pei[2], Feiyue Shang[3]

[1] School of Economics and Management, China University of Petroleum (Beijing), Beijing, 100000

[2] School of Management, Soochow University, Suzhou, Jiangsu, 215021

[3] College of Economics and Management, Shandong University of Science and Technology, Qingdao, Shandong, 266590

Keywords: deep learning; neural network; stock price; LSTM; CNN

Abstract: In this paper, we use the deep learning method to predict the rising and falling direction of the Shanghai and Shenzhen 300 index from 2012 to 2019. After feature extraction is carried out by using convolution neural network and short-term memory model on multiple time scales, the final prediction results are obtained by splicing the feature matrices on different time scales. Compared with other models, the experimental results show that the multi-time-scale CNN-LSTM model proposed in this paper can effectively improve the effect of forecasting the rise and fall of CSI 300 index and make a profit in the trading back test. The research content of this paper enriches the methods of financial time series data analysis, which can not only provide decision-making reference for investors, but also help to enhance the understanding of the laws of the financial market.

1. Introduction

Affected by many factors, stock price is an extremely complex dynamic system, which has the characteristics of non-linearity, non-stationarity, low signal-to-noise ratio and long memory. The classical financial time series forecasting model needs to assume that the data meet some specific requirements, but the analytical equation with parameters is difficult to accurately describe such a complex dynamic system, so stock price forecasting has been difficult to achieve good results.[1] Since artificial intelligence program AlphaGo defeated world go champion Lee se-dol in 2016, its deep learning method has been successfully applied in image recognition, speech recognition, self-driving and other fields. For the financial field, artificial intelligence has been more and more studied and applied in both academia and industry, together with big data, block chain and other technical means as an important technology driver of the current financial science and technology strategy.[2]

Deep learning is a machine learning method based on multi-layer neural network, which is composed of neural network output layer and input layer and a series of stacked hidden layers.[3] The highly abstract and complex features of learning objects are extracted through layer-by-layer transmission, and the characteristic matrix is used as the representation of data to obtain highly complex nonlinear functions between input data and output data. [4]Thus ultimately improve the accuracy of classification or prediction. Based on the advantage of deep learning in dealing with high-dimensional, nonlinear and non-stationary data structures, the question to be studied in this paper is whether the method of deep learning can effectively model the financial time series data and improve the prediction accuracy of stock prices. In-depth study of these issues can not only provide decision-making reference for investors, but also enrich financial time series data processing and research methods, and help to improve the ability of supervision and early warning of financial market risks.[5]

In view of the above problems, taking the CSI 300 index as an example, this paper combines the long-and short-term memory model (LSTM), which is widely used in the field of deep learning, with the convolution neural network (CNN), so as to make use of its advantage of feature extraction at the same time, and applies the above compound depth neural network to the daily K-line time scale and 15-minute K-line time scale data respectively. The input data on the two time scales are mapped to the high-dimensional feature space respectively, and then the characteristic matrix is spliced to generate the final result of the rise and fall prediction. After comparing the super-parameters of different network structures, the empirical results show that the prediction effect of CNN-LSTM composite model is better than that of traditional shallow neural network and single model depth neural network. The prediction effect of compound depth neural network can be further improved by using multi-time-scale data fusion prediction.[6]

2. Research methods and models

2.1. Convolution neural network

The convolution neural network, which draws lessons from the autogenous visual system, is a kind of incompletely connected deep neural network with convolution structure. The neurons in each layer are locally connected to realize the extraction and transformation of the hierarchical features of the input. The neurons with the same connection weight are connected to different regions of the upper neural network. Furthermore, the network parameters are simplified and the network has a certain degree of stability of displacement, scale, scaling and nonlinear deformation.

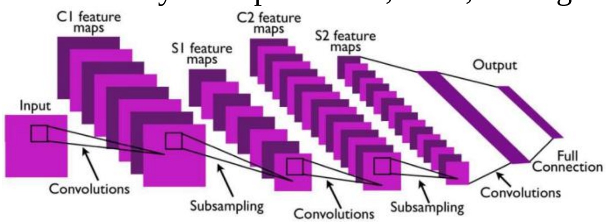

Figure 1 Schematic diagram of convolution neural network structure

The structure of convolution neural network is shown in figure 1. In addition to input layer and output layer, convolution layer, pooling layer and full connection layer are also included. Among them, the convolution layer carries out feature extraction through the convolution check input space shared by weights, and the specific process can be expressed as follows:

$$h_{j,k} = \sigma(b + \sum_{l=0}^{L} \sum_{m=0}^{M} w_{l,m} \alpha_{j+l,k+m})$$

After the feature extraction of the convolution layer, the next step of the convolution neural network is pooling. The pool layer usually appears in pairs with the convolution layer, and its function is to downsample the output obtained by the convolution layer, simplify the information and enhance the generalization ability of the model. After the feature extraction and abstraction of several groups of convolution layer and pooling layer, the previous output is weighted and classified by the traditional full connection layer.

2.2. Long-and short-term memory model

In the traditional cyclic neural network, when the time-based back propagation algorithm (BPTT) is used for training, the gradient disappears with the increase of the number of recursive layers, which will seriously affect the effect of the model, while the long-term memory model (LSTM) optimizes this from the structure of neural units. Similar to the traditional cyclic neural network, LSTM also has a chain module structure, but the composition structure of repetitive modules is different. As shown in figure 2, LSTM realizes the function of selective forgetting and memorizing information by designing three thresholds of forgetting gate (Forget Gate), input gate (Input Gate) and output (Output Gate) in the memory unit, and adds cell state (Ct) to save long-term information in the sequence, and the function of selective forgetting and memorizing information is reflected in the modification of cell state.

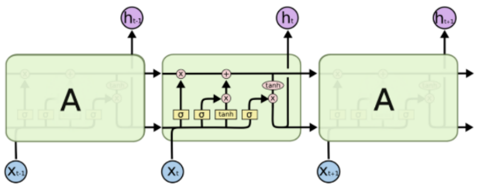

Figure 2 Schematic diagram of LSTM structure

3. Empirical analysis

3.1. Data sources and preprocessing

This paper forecasts the rising and falling direction of the CSI 300 Index in the next trading day, and selects the data of the opening price, the highest price, the lowest price, the closing price and the trading volume of the daily K line and the 15-minute K line from January 1, 2012 to September 30, 2019, compared with the previous trading day or 15 minutes. The data sources are wind database and stock trading access letter software. Among them, there are 1883 trading days on the daily K-line time scale and 30129 minute K-line on the 15-minute K-line time scale. The time step used on the daily K-line time scale is 20, that is, the price information of the first 20 trading days is used to predict the rising and falling direction of the stock price of the next day, and the time step used on the 15-minute K-line time scale is 16. That is, the price information of 16 15-minute K lines of the previous trading day is used to predict the rising and falling direction of the stock price the following day. Therefore, there are 1862 available samples in this experiment. If 80% of the training set is used as the training set and 20% as the test set, the number of samples in the training set is 1489 and the number of samples in the test set is 373. For the obtained stock data, this paper uses the Z-score method to preprocess the data. The Z-score method is a centralization method, which standardizes the original data based on the mean and standard deviation of the original data. The mathematical expression is as follows:

$$\bar{x_t} = \frac{x_t - \text{mean}(x)}{\text{std}(x)}$$

Among them, xt is the original stock price data, mean (x) is the mean value of stock data, std (x) is the standard deviation of stock data, $\bar{x}$is the standardized stock price data.

3.2. Evaluation index

In view of the fact that this paper is to predict the rise and fall of the stock index, we select the accuracy (Accuracy), accuracy (Precision), recall rate (Recall), F1 value (F1score) and area under the curve (AUC) which are often used in the classification task as the evaluation index of the prediction effect. Among them, the accuracy reflects the correct sample proportion in the total sample; the accuracy reflects the actual rising sample proportion in the predicted rising sample; the recall rate reflects the predicted rising sample proportion in the actual rising sample; the F1 value is the harmonic average of accuracy and recall rate. AUC is the area surrounded by the coordinate axis under the receiver operating characteristic curve (ROC), which represents the probability that a positive sample and a negative sample are randomly selected, and the predicted value of the positive sample is greater than that of the negative sample. This index can avoid the evaluation distortion caused by the imbalance of samples.

3.3. Benchmark model

3.3.1. BP neural network

When applying neural network to learning tasks, a very important task is to find a suitable network structure, that is, hyperparameters such as the number of hidden layers and the number of neurons in each layer. the selection of different hyperparameters has a great impact on the results of machine learning tasks. however, at present, there is no theoretical basis for the selection of hyperparameters in academic circles. In practical applications, different network structures are usually tried repeatedly, and a set of super parameters with the best results are selected. In this section, the ReLU activation function is used in the hidden layer of the BP neural network, and the sigmoid activation function is used in the output layer. In the iterative optimization process, the Adam method is used to try to conceal the hyperparameters that the number of layers is two, three, four and five, and the number of neurons in the hidden layer is 50,100,150,200. thus a more suitable BP neural network structure is found. Table 1 shows the accuracy of using BP neural networks with different structures to predict the CSI 3000.The optimal combination is composed of three hidden layers and the number of neurons in each hidden layer is 100. the highest prediction accuracy of the test set is 54.67%.

Table 1 Prediction accuracy of BP Neural Networks with different Network structures

	2 hidden layers	3 hidden layers	4 hidden layers	5 hidden layers
50 neurons	53.33%	53.87%	53.33%	53.33%
100 neurons	53.87%	54.67%	53.33%	53.07%
150 neurons	53.60%	53.07%	54.13%	53.87%
200 neurons	53.60%	52.80%	54.40%	54.40%

Table 2 Prediction accuracy of BP Neural Networks with different Network structures

	Single cycle layer	2 circular layers	3 circular layers
200 neurons	54.42%	56.30%	55.23%
150 neurons	56.57%	55.77%	55.23%
100 neurons	57.11%	56.84%	54.16%
50 neurons	56.03%	55.50%	54.16%

3.3.2. LSTM

In the LSTM neural network, the main network structures that affect the prediction results include the time step, the depth of the loop layer and the number of neurons in each loop layer. In this section, the Adam optimization method is used in the LSTM model, and the learning rate is 0.001. The activation function of the hidden layer is the Re activation LU function, and the activation function of the output layer uses the sigmoid function. By comparing the number of different cycle layers and the number of neurons under the 20-day time step, the prediction results of the test set are shown in Table 2. Under the network structure of 100 neurons in a single cycle layer, the prediction accuracy can reach 57.64%.

3.3.3. Convolution neural network

At present, the commonly used convolution neural network is mainly two-dimensional convolution neural network, which can effectively extract spatial relations and features from data through two-dimensional convolution kernel. however, there is no meaningful two-dimensional spatial relationship in the general storage structure of financial time series data. The main difference between one-dimensional convolution neural network and two-dimensional convolution neural network lies in the size and window sliding mode of convolution kernel. In two-dimensional convolution neural network, the convolution window will move horizontally and longitudinally to extract the input data memory features. In the one-dimensional convolution neural network, by setting the size of the convolution kernel to be the same width as the size of the input data, the convolution window only slides longitudinally. This network structure is mainly used to extract the translation features of input data in a single spatial direction, so it may have better applicability for financial time series data. Therefore, this paper uses two-dimensional convolution neural network and one-dimensional convolution neural network to predict and compare under a variety of network structures. As shown in Table 3, the size of the convolution kernel is shown in parentheses.

Table 3 Prediction accuracy of CNN models with different Network structures

2D CNN (2x2)	2D CNN (3x3)	2D CNN (4x4)	2D CNN (5x5)	2D CNN (6x6)
54.42%	53.89%	55.76%	53.62%	56.57%

3.4. Multi-time scale compound depth neural network

3.4.1. Model building

In order to better apply the deep learning method to predict the stock price, this paper combines the two deep learning models to predict the stock price by using the feature advantage of simultaneously extracting the long-and short-term memory model and the convolution neural network. in order to better process and analyze the financial time series data, and further through the daily K-line and 15-minute K-line data to extract features for feature matrix stitching. Thus, we can extract the medium-and long-term information and short-term information from stock data at the same time, and make use of the characteristics of different time scales to achieve better prediction results.

The multi-time-scale compound depth neural network used in this paper includes two models, namely, multi-time-scale CNN-LSTM neural network and multi-time-scale LSTM-CNN neural network. In the comparative analysis of the previous benchmark model, the prediction performance of one-dimensional convolution neural network and LSTM model with 100 neurons in single cycle layer is better. Taking the multi-time-scale CNN-LSTM neural network composed of these two models as an example, the specific data processing and feature learning process are introduced in detail.The model first constructs a model Model1 that receives daily K-line time-scale data and a model Model2 that receives 15-minute K-line time-scale data. In Model1 and Mod inverse el2, the input data will first pass through a convolution layer, and the first-level features of the input data will be extracted by using one-dimensional convolution kernel, and then the output data of the convolution layer will be input into the loop layer to extract the second-level features. Then the secondary features obtained under the two time scales are merged and input to the full connection layer, and then processed by the Dropout layer and input to the output layer to get the final prediction results. The specific data processing and feature learning process are as follows:

(1) Input layer. The input of the daily K-line time scale is the change range of the opening price, the highest price, the lowest price, the closing price and the trading volume processed by z-score compared with the previous trading day. Each sample contains the above data for 20 trading days, so the single sample of the input data is a matrix of 20 rows and 5 columns. The input of the 15-minute K-line time scale is the change of the opening price, the highest price, the lowest price, the closing price and the trading volume processed by z-score compared with the previous 15 minutes. Each sample contains the above data of 16 15-minute K-lines, so the single sample of the input data is a matrix of 16 rows and 5 columns.

(2) Convolution layer. The input data of two time scales are convoluted with one-dimensional convolution kernels of size 5, and the number of convolution kernels is 16, that is, the input space of each convolution kernel slides down according to the window size of 1: 5 to extract features. A total of 16 different features are extracted. The output of a single sample in this layer on two time scales is a matrix with dimensions of (20) and (16) respectively.

(3) Cycle layer. The output of the convolution layer is taken as the input of this layer, that is, the input data is a data structure with a time step length of 20 and a feature dimension of 16 on two time scales, and a data structure with a time step length of 16 and a feature dimension of 16. A LSTM hidden layer containing 100 neurons is used to process the input data respectively, and the output of a single sample in this layer is a dimension of (1) on two time scales. 100).

(4) Full connection layer. After merging the output of the loop layer on two time scales, the input data structure of a single sample in the full connection layer is a matrix of (1200), and then the output data of the previous layer is processed by a full connection layer with 20 neurons. The output of a single sample in this layer is a matrix with a dimension of (1).

(5) Dropout layer. The function of the Dropout layer is to randomly freeze some neurons in the hidden layer from the neural network during each iterative training, and reduce the correlation between the neurons and the complexity of the model in this way, so as to achieve the effect of regularization. A Dropout ratio of 0.2 is used in this model.

(6) Output layer. The output of the upper layer is processed by the full connection layer which contains only one neuron, and the final output, that is, the prediction result of the model, is obtained. The output is the probability that the CSI 300 index will rise in the next trading day. When the rising probability exceeds 0.5, the predicted result is considered to be rising, otherwise it is regarded as falling.

The prediction accuracy of multi-time-scale composite neural networks under different network structures is shown in Table 4, in which N1, N2 and N3 represent the number of neurons as 50, 100 and 200 respectively. It can be seen that the prediction accuracy of multi-time-scale CNN-LSTM is higher than that of multi-time-scale LSTM-CNN, under various model structures, which indicates that multi-time-scale CNN-LSTM is more suitable for the analysis and prediction of financial time series data. When using a single-layer cyclic layer with 100 neurons, the prediction accuracy of multi-time-scale CNN-LSTM is 60.86%, which is better than other network structures. The ROC curve is shown in figure 4, the Abscissa is the false positive class ratio (FPR), the vertical axis is the true class rate (TPR), and the area under the curve, that is, the AUC value is 0.612. The above results show that the prediction effect of the model is good and balanced in both cases. The forecast object of this paper is the rising and falling direction of the CSI 300 index in the next trading day. In order to more fully reflect the forecasting effect of the multi-time-scale CNN-LSTM on the stock price data, this paper also uses this model to predict the CSI 300 index.

Table 4 Prediction accuracy of multi-time-scale compound neural network

	Single layer cycle layer			Two-layer cycle layer		
	N1	N2	N3	N1	N2	N3
Multi-scale CNN-LSTM	60.32%	60.86%	58.71%	58.45%	59.51%	58.18%
Multi-scale LSTM-CNN	57.11%	56.57%	57.91%	57.91%	58.45%	57.91%

In order to better measure the prediction effect of multi-time-scale CNN-LSTM neural network, the evaluation index of its prediction results is compared with BP neural network, LSTM neural network, CNN neural network and compound depth neural network. As shown in Table 5, the prediction effect of depth neural network is better than that of BP neural network, the prediction effect of compound depth neural network is better than that of single model, and the prediction effect can be further improved by fusing multi-time-scale data on compound depth neural network. The prediction accuracy of multi-time-scale CNN-LSTM model is 6.19% higher than that of shallow BP neural network, and 3.75% higher than that of LSTM model, which is the most widely used in financial time series data. Other evaluation indicators have also been significantly improved.

Table 5 Comparison of prediction effects of each model

BP	Accuracy	Precision	Precision	Precision	Precision
2D-CNN	54.67%	55.65%	55.65%	55.65%	55.65%
1D-CNN	55.76%	55.62%	55.62%	55.62%	55.62%
LSTM	56.57%	56.59%	56.59%	56.59%	56.59%
	57.11%	57.65%	57.65%	57.65%	57.65%
LSTM-CNN	56.84%	57.45%	57.45%	57.45%	57.45%
CNN-LSTM	57.91%	57.92%	57.92%	57.92%	57.92%
Multi-scale LSTM-CNN	59.25%	59.60%	59.60%	59.60%	59.60%
Multi-scale CNN-LSTM	60.86%	61.50%	61.50%	61.50%	61.50%

4. Conclusion

Deep learning has great advantages in dealing with high-dimensional non-linear systems. As a non-parametric analysis method, deep learning does not need to assume that the analysis object conforms to a specific equation structure or distribution function, so it has good applicability and research prospect for financial time series data. The research of this paper enriches the methods of financial time series data analysis. Different from the traditional econometric methods, the deep learning method pays more attention to the model structure, feature extraction and fitting effect.

References

[1] SCHMIDHUBER J. Deep Learning in Neural Networks: An Overview[J]. Neural Networks, 2015, 61: 85-117.

[2] NAJAFABADI M M,VILLANUSTRE F,KHOSH- GOFTAAR T M,et al. Deep Learning Applications and Challenges in Big Data Analytics[J]. Journal of Big Data, 2015, 2 (1): 1-21.

[3] FAMA E F. The Behavior of Stock- Market Prices[J]. The Journal of Business, 1965, 38 (1): 34-105.

[4] FAMA E F. Efficient Capital Markets: A Review of The- ory and Empirical Work[J]. The Journal of Finance, 1970, 25 (2): 383-417.

[5] GROSSMAN S J. On the Efficiency of Competitive Stock Markets where Traders Have Diverse Information [J]. Jour- nal of Finance, 1976, 31 (2): 573-585.

[6] GROSSMAN S J. An Introduction to the Theory of Ratio- nal Expectations under Asymmetric Information [J]. Review of Economic Studies, 1981, 48 (4): 541-559.