Research Based on Stock Predicting Model of Neural Networks Ensemble Learning

Xie Qi[1,a], Cheng Gengguo[1], Xu Xu[2] and Zhao Zixuan[3]

[1]School of information science and engineering, Wuhan University of Science and Technology, Wuhan, Hubei, 430065, China
[2]Ecole Supérieure de Commerce, Grenoble Ecole de Management, Grenoble, 38000, France
[3]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, China

Abstract. Financial time series is always one of the focus of financial market analysis and research. In recent years, with the rapid development of artificial intelligence, machine learning and financial market are more and more closely linked. Artificial neural network is usually used to analyze and predict financial time series. Based on deep learning, six layer long short-term memory neural networks were constructed. Eight long short-term memory neural networks were combined with Bagging method in ensemble learning and predicting model of neural networks ensemble learning was used in Chinese Stock Market. The experiment tested Shanghai Composite Index, Shenzhen Composite Index, Shanghai Stock Exchange 50 Index, Shanghai-Shenzhen 300 Index, Medium and Small Plate Index and Gem Index during the period from January 4, 2012 to December 29, 2017. For long short-term memory neural network ensemble learning model, its accuracy is 58.5%, precision is 58.33%, recall is 73.5%, F1 value is 64.5%, and AUC value is 57.67%, which are better than those of multilayer long short-term memory neural network model and reflect a good prediction outcome.

1 Introduction

In recent years, the theory of stock market prediction has become more and more matured. From the random walk theory and the effective market hypothesis to the quantitative investment, the prediction of the time series and the financial trading are more closely linked with machine learning. Machine learning has one function named self-perfection, and with the development of artificial intelligence, machine learning has been widely used in financial field. This paper proposes a financial time series prediction model based on ensemble learning of long short-term memory neural network. The financial time series data set used in this paper is composed of 6 indices of the Chinese stock market. Each index converts the data from two-dimensional data into three-dimensional data and normalizes each of the three-dimensional data. This neural network consists of 6 layers, including four LSTM layers, one connected layer and one activated layer. The activation function of the LSTM layer is tanh, and the deep neural network training process uses Adam algorithm as an optimization algorithm. The multi-layer long short-term memory neural network is used as the base learner, and the data set is divided into several sub data sets to train those classifiers. The final classification result is formed by integrating the classification results of the base learners and it is the estimation of next trading day.

2 Background

The prediction of the financial time series is always a hot spot in the financial circle and academia. In 1959, Osborne[1] proposed random walk theory based on Brownian movement. The theory holds that the change of stock price is the reaction of the market to random event, and it is similar to Brownian movement, which has the characteristics of random walk meaning that change path has no rules to be followed. Therefore, the theory holds that the fluctuation of stock prices is unpredictable. In 1970, Fama[2], the Nobel prize winning economist, proposed efficient market hypothesis, which holds that and that all valuable information has been reflected in the stock price trend in time, accurately and fully in the securities market with sound legal system, good function, high transparency and efficient competition based on the assumption that all investors are rational. Accordingly, price is unpredictable in the long term, and no analytical method can predict the trend of stock price effectively.

However, some scholars held the opposite views on random walk theory and efficient market hypothesis. In 1999, Andrew and Mackinlay [3] proposed non-random walk theory, holding that the change of stock price does not follow the random walk theory. Non-random walk theory uses the economic model to handle the historical data, sums up the law of the stock price volatility, and gets the rate of return higher than the market overall level according to the law. Therefore, the theory holds that stock prices can be predicted.

ᵃ Corresponding author: 85169023@qq.com

In 1971, Barclays Investment Management Inc. released the first passive managed index fund in the world, which marked the beginning of quantitative investment. Quantitative investment has become an important way of investment in the US market. The proportion of US quantitative investment rose to over 30% in 2009 [4], and about 85% of the US stock market transactions were completed through algorithmic trading [5]. In 1988, James Simons, a data scientist, established Medallion fund, which used a quantitative model to invest, and has achieved return of an average of 34% per year since its establishment. Non-random walk theory and the practical application of quantitative investment prove the feasibility of stock price prediction.

Machine learning improves the performance itself by means of computation and experience[6]. Some scholars use various machine learning algorithms to sum up the historical rules of financial data in order to improve the accuracy of investment decisions. In 1999, Allen used genetic algorithm to derive the technical trading rule from the historical price of the US stock market [7]. In 2013, Kim tested the feasibility of SVM in financial prediction by comparing SVM and neural network. [8] In 2016, Khaidem used the random forest algorithm to predict stock returns to minimize the risk of stock investment [9].

Artificial neural network is a hot spot in artificial intelligence. In 1986, Rumelhart first proposed back propagation neural network [10]. In 1988, White used neural network to model to explain the nonlinear rule of asset price change and forecast the daily price change of IBM stock [11]. In 1997, Hochreiter and Schmidhuber proposed the Long short-term memory (LSTM) [12]. LSTM model avoids the negative influence of vanishing gradient and exploding gradient through the control of input gate, output gate and forget gate. LSTM model is one kind of the recurrent neural networks, which can theoretically store infinite number of time information. In 2006, Professor Hinton, professor of University of Toronto in Canada and leading authority in the field of machine learning, and his student Salakhutdinov proposed using Multi-Layer neural network to reduce the data dimension in the top academic journal Science [13]. This article provided two main pieces of information: (1) the artificial neural network with multiple hidden layers has excellent feature learning ability, and the acquired features explain well for data, which is beneficial to visualization or classification; (2) the difficulty in training of deep neural networks can be effectively overcome by "layer by layer initialization". [14]. Multi hidden layer neural network can extract more excellent features. Deep learning does well in speech recognition and image recognition.

Some scholars apply deep learning to stock price forecasting and quantitative investment. In 2015, Chen used the LSTM model to predict stock price of Chinese market [15]. In 2015, Ding and others proposed the use of deep learning methods, which is to extract features from the news text and use convolution neural network to predict the short-term and long-term trend, to predict event driven stock price changes [16]. In 2016, Jia validated the validity of the LSTM model in predicting the stock price trend [17]. In 2017, Nelson and others used the LSTM model to predict the future stock market trend based on historical prices and technical indicators, and compared it with other machine learning methods [18].

3 Data and Pre-processing

The data set consists of 6 indexes, namely Shanghai Composite Index, Shenzhen Composite Index, Shanghai Stock Exchange 50 Index, Shanghai-Shenzhen 300 Index, Medium and Small Plate Index and Gem Index. Data for each index covers from January 4, 2012 to December 29, 2017, so that there is a total of 1458 pieces. Each piece of data contains thirteen characteristics, including opening price, closing price, highest price, lowest price, volume, MACD and KDJ, and one target attributes, which represents its rise or fall in next trading day. The target attribute will be 1 if the index in next trading day rise and will be 0 if the index in next trading day fall.

The LSTM model can store long time information in time series, so it can be used to handle and predict financial time series. The LSTM model is different from other machine learning models in data format. The traditional machine learning algorithm, such as the logistic regression algorithm, is one-dimensional data and the data set of it is two-dimensional, while each piece of data from the LSTM algorithm is from the past so that it is two-dimensional and the data set of this is three-dimensional. Original two-dimensional time series data sets are transformed into three-dimensional time series data by using LSTM to train and predict financial time series. The meaning of each two-dimensional data in the three-dimensional data is the stock index change in the past period. The target attribute corresponding to each two-dimensional array is the next trading day after a continuous period of time. Each two-dimensional array and its corresponding target data form a complete panel data. Multiple panel data sets up a data set.

Different indicators often have different dimensions and units, which have negative impact on the result of data analysis. In order to eliminate the impact of dimensions among indicators, standardization is needed to make data indicators comparable. After standardization, original data is in the same order of magnitude, which is suitable for comprehensive comparative evaluation. This paper adopts Min-Max Scaling method, which is also called Min-Max normalization, and its formula is shown in formula (1).

$$y = \frac{x - MinValue}{MaxValue - MinValue} \tag{1}$$

The method realizes the equal scaling of the original data, in which y represents the normalized data, x represents the original data, and *MaxValue* and *MinValue* respectively represent the maximum and minimum value of the original data set.

After three-dimensional and normalization, according to the order of financial time series, the data are divided into three parts, training set data, validation set data and

test set data. 80% of data set is used as a training set to train classifiers, 10% is used as a validation set to verify the number of classifiers which has the best training effect, and other 10% of is used as the test set to test the outcome.

4 Equations and mathematics

4.1. Long short-term memory neural networks

Deep neural network is a machine learning model for processing complex nonlinear mapping between input and output. Deep neural network has excellent feature learning ability and big data processing advantages. Deep neural networks have problems of vanishing gradient and exploding gradient. The gradient of the error function of neural network is propagated back and it will be scaled. Recurrent neural network is a kind of deep neural network and its gradient of recurrent neural network grows exponentially or decays over time.

Long Short-Term Memory (LSTM) is a kind of recurrent neural network. LSTM controls information through structures of three gates to solve the problem of vanishing gradient and exploding gradient produced by recurrent neural network. In the process of information transmission, LSTM can retain information before. LSTM is suitable for processing and predicting important events with very long intervals and delays in time series. In stock time series, the former information is closely related to the subsequent information as it is a factor that affects the subsequent information. LSTM is suitable for the analysis and prediction of stock time series.

Deep neural network is a very powerful machine learning system. Overfitting is a serious problem in deep neural networks. Large scale deep neural network system is slow and difficult to solve the problem of overfitting. In order to solve this problem, in 2014, Srivastava and Hinton proposed Dropout method to prevent neural networks from overfitting [19]. The Dropout method is to randomly discard cells in the process of training neural network, so that the weights of neurons can be prevented from excessive adjustment. This method avoids overfitting problem and is more effective than other regularization methods.

Adam (Adaptive Moment Estimation) is another adaptive learning rate method. Before the Adam algorithm, the training neural network optimization algorithm uses the gradient descent optimization or the random gradient descent optimization. In the process of training the data, they minimize the loss function by optimizing the loss function, and iteratively updates the weight of the neural network. But these algorithms consume a lot of time and computing resources. In 2014, Kingma and Jimmy Ba proposed the Adam algorithm, which dynamically adjusted the learning rate of each parameter by using the first order and second order moment estimation of gradient. The Adam algorithm has high computational efficiency and low storage requirements. The gradient diagonal scaling of Adam algorithm is invariable, so it is very suitable for solving

problems with large-scale data or parameter. The algorithm is also applicable to solve the unsteady problem of large noise and sparse gradient. Experimental results show that the Adam algorithm is better than other optimization methods [20].

The multilayer long short-term memory neural network proposed in this paper consists of six layers: the first layer consists of 64 neurons and is imported as 13 characteristics, the second layer consists of 512 neurons, the third layer consists of 128 neurons and the fourth layer consists of 64 LSTM elements. Those four layers are activated by tanh function, shown in the formula (2).

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

In the first four layers, Dropout method is used after every last layer to reduce the overfitting and promote the generalization of the model. The fifth layer is connected layer and the sixth layer is activated layer. The activated function is sigmoid function, shown in the formula (3).

$$S(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

In formula (2) and (3), x represents neural network data while $f(x)$ represents the outcome after activation.

Because the data set for prediction is a binary classification problem, the loss function of the multilayer long short-term memory neural network is based on the cross-entropy loss function. In the training process, the Adam algorithm is used as an optimization algorithm. The model of the multilayer long short-term memory neural network is shown in Figure 1.
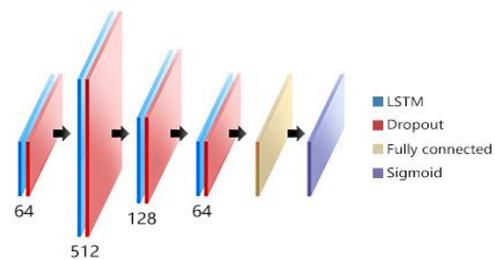

Figure 1. Multilayer long short-term memory neural networks

4.2 Long short-term memory neural network ensemble learning

Ensemble learning is to form a strong classifier by several weak classifier to improve the performance. Ensemble learning can be divided into three categories, which are Bagging, Boosting and Stacking. Bagging is the abbreviation of Bootstrap aggregating. In 1996, Leo Breiman proposed the Bagging algorithm [21]. Given one training set, the Bagging algorithm uses self-help sampling method to select N subset training sets randomly [22]. Each sub training set is trained to generate a model and a total of N models are obtained. The final result of the test is obtained by N model. If it is a classification problem, the results of N test are voted, and if it is the regression problem, the results is the mean value of the N test results.

In 1990, Hansen proposed ensemble learning of neural networks. Each neural network was regarded as a weak classifier, and the prediction results of multiple neural networks were combined into the final output under crossover.

The method of verification can optimize the parameters of neural network. The experimental results of Hansen show that ensemble learning of neural network can improve the generalization of neural network [23].

Long short-term memory neural network ensemble, which is different from the traditional ensemble learning neural network, is to replace traditional neural network with the long short-term memory neural network. Using self-help sampling method, the training set is divided into several sub training sets, and each sub training is trained through LSTM algorithm to train a weak classifier. The Bagging algorithm based on ensemble learning is used to vote on the classification results of the weak classifiers, and the voting categories which account for more than half are the final classification results.

4.3 Long short-term memory neural network ensemble learning

The prediction model based on long short-term memory neural network ensemble learning is divided into four parts. The first step is to pre-process the data of the financial time series, convert the data set from the two-dimensional data set to the three-dimensional data set, normalize the three-dimensional data set, and finally divide data set into training data set, validation set and test data sets according to the time series order. The second step uses the Bagging algorithm to divide the test data set into a number of sub test data sets, and trains each of the sub test data sets to train a number of long short-term memory neural networks. The third step is to predict based on the test data using the long memory neural network, calculate the best classifier number out. The fourth step is to use the testing set to test each long short-term memory neural networks and use the Bagging algorithm to set the multiple test result sets as the final prediction results. The fifth step is to compare the real result to the forecast result.

5 Experiment

The experiment is divided into three parts. The first part tests the predictive performance of the multi-layer memory neural network. The second part tests the best number of classifiers. The third part tests are based on the prediction of the integrated learning of the long and short memory neural network. Compare the test results of these two parts.

The experiment tested Shanghai Composite Index, Shenzhen Composite Index, Shanghai Stock Exchange 50 Index, Shanghai-Shenzhen 300 Index, Medium and Small Plate Index and Gem Index during the period from January 4, 2012 to December 29, 2017. Each index tested for 1458 data. Python 3 was used to do the

experiment and keras was used to build the neural network structure.

Based on the real results and the prediction results of the sample, the sample are divided into four categories.

(1)TruePositive TP, the number the sample is true while the prediction result is positive.

(2)FalseNegative FN, the number the sample is false while the prediction result is negative.

(3)FalsePositive FP, the number the sample is false while the prediction result is positive.

(4)TrueNegative TN, the number the sample is true while the prediction result is negative.

The evaluation of the classification performance and classification algorithm uses the following five indicators: Accuracy, Precision, Recall, F1, which are respectively calculated under formula (4), (5), (6) and (7), and AUC. AUC is the area under the ROC curve, and the range of AUC is between 0.5 and 1. The larger the AUC, the better the classification effect of the algorithm.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

The first part of the experiment tested the predicting performance of multilayer long short-memory neural network. Using Shanghai Composite Index, Shenzhen Composite Index, Shanghai Stock Exchange 50 Index, Shanghai-Shenzhen 300 Index, Medium and Small Plate Index and Gem Index, the multilayer long short-term memory neural network was trained respectively, and the after-training neural network was used to test the test set of 6 indexes. The accuracy, precision, recall rate, F1 value and AUC value of the 6 indexes were calculated.

In the training process, each layer of the Dropout parameter is set to 0.2, which means 20% neurons in each layer are dropout. The Batch value is set to 30, and the number of samples contained in each batch of gradient descent training is 30. The epochs value is set to 30, which mean it will stop when it trains 30 times. The validation_split value is set to 0.3, which means 30% of the training data is the validation data.

The outcome of testing was shown in table 1. The average accuracy is 47.33%, the average precision is 49.83%, the average recall rate is 63.5%, the average F1 value is 55.5% and the average AUC value is 48.17%.

Table 1. Predicting performance of multilayer long short-term memory neural networks

	Accuracy	Precision	Recall	F1	AUC
Shanghai Composite Index	0.42	0.48	0.51	0.5	0.41
Shenzhen Composit	0.4	0.49	0.75	0.59	0.49

e Index					
Shanghai Stock Exchange 50 Index	0.52	0.53	0.68	0.59	0.51
Shanghai - Shenzhen 300 Index	0.5	0.57	0.6	0.59	0.48
Medium and Small Plate Index	0.49	0.49	0.75	0.59	0.49
Gem Index	0.51	0.43	0.52	0.47	0.51

The second part of the experiment tested the number of classifiers under the best training circumstance. A validation set is used to verify the prediction effect of classifiers from 2 to 10 classifiers respectively. Each classifier tested 6 indexes, each index is tested for 10 times, and the results use average value. The F1 value considers both the precision and recall, so the number of classifiers corresponding to the highest F1 value is chosen as the optimal parameter of ensemble learning. As shown in Figure 1, the F1 value is the highest when the number of classifiers is 8, so 8 becomes the number of ensemble learning classifiers.
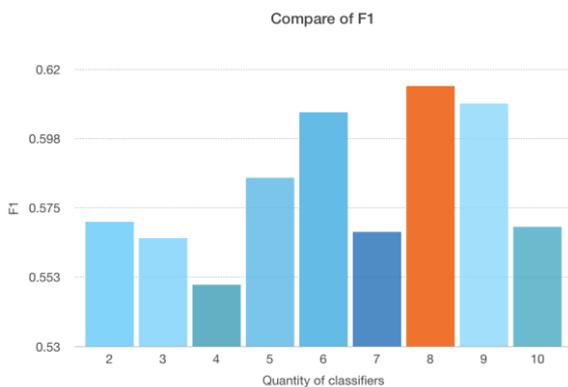

Figure 2. Comparison among different F1 value under different numbers of classifiers

The third part of the experiment tested the prediction performance of ensemble learning based on long short-term memory neural network. The training sets using the Shanghai index, the Shenzhen index, the Shanghai 50 index, the Shanghai and Shenzhen 300 index, the SME board index and the gem index are generated by the self-help sampling method to generate 8 sets and use 8 new training set training models. Test set was used to test ensemble learning model.

The outcome of testing was shown in table 2. The average accuracy is 58.5%, the average precision is 58.33%, the average recall rate is 73.5%, the average F1 value is 64.5% and the average AUC value is 57.67%.

Table 2. Predicting performance of long short-term memory neural networks ensemble learning

	Accuracy	Precision	Recall	F1	AUC
Shanghai Composite Index	0.6	0.61	0.83	0.7	0.57
Shenzhen Composite Index	0.58	0.55	0.83	0.65	0.58
Shanghai Stock Exchange 50 Index	0.58	0.62	0.63	0.63	0.57
Shanghai-Shenzhen 300 Index	0.62	0.66	0.72	0.69	0.59
Medium and Small Plate Index	0.58	0.56	0.64	0.6	0.58
Gem Index	0.55	0.5	0.76	0.6	0.57

Compared with table 1 and table 2, it is found that compared with the outcome of multilayer long short-term memory neural network, the accuracy of long short-term memory neural network ensemble learning increases by 11.17%, the precision increases by 8.5%, the recall rate increased by 10%, the F1 value increased by 9%, and the AUC value increased by 9.5%. Experiments show that long short-term memory neural network ensemble learning has higher predicting performance than multilayer long short-term memory neural networks.

6 Conclusion

In this paper, a financial time series prediction model based on the ensemble learning of long short-term memory neural network is proposed. After the model is used to predict 6 main indexes of Chinese stock market and 5 characteristics, such as accuracy and accuracy, the following conclusions are drawn:

(1) The outcome of experiment is that the average accuracy is 58.5%, the average precision is 58.33%, the average recall rate is 73.5%, the average F1 value is 64.5% and the average AUC value is 57.67%. The accuracy and AUC value of each index are predicted to be higher than 50%. The results of experiment show that the prediction results are superior to the random test. The changes in the 6 indices of the Chinese stock market do not follow the random walk theory, and the 6 index rises and falls can be predicted.

(2) Compared with the outcome of multilayer long short-term memory neural network, the accuracy of long short-term memory neural network ensemble learning increases by 11.17%, the precision increases by 8.5%, the recall rate increased by 10%, the F1 value increased by 9%, and the AUC value increased by 9.5%. Experiments show that long short-term memory neural network ensemble learning has higher predicting

performance than multilayer long short-term memory neural networks.

References

1. Osborne M F M. Brownian Motion in the Stock Market[J]. Operations Research, 1959, 7(2):145-173.

2. Malkiel B G, Fama E F. EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*[J]. Journal of Finance, 1970, 25(2):383-417.

3. Lo A W, Mackinlay A C. A Non-Random Walk Down Wall Street:[M]. PRINCETON UNIVERSITY PRESS, 1999.

4. Ding Peng, Quantitative Investment: Strategy and Technology[M]. PHEI, 2014.

5. M. Glantz, R. Kissell, Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era[M], Academic Press, 2013.

6. Zhou Zhihua, Machine learing[M]. Tsinghua University Press, 2016.

7. Allen F, Karjalainen R. Using genetic algorithms to find technical trading rules 1[J]. Journal of Financial Economics, 1999, 51(2):245-271.

8. Kim K J. Financial time series forecasting using support vector machines[J]. Neurocomputing, 2003, 55(1):307-319.

9. Khaidem L, Saha S, Dey S R. Predicting the direction of stock market prices using random forest[J]. 2016.

10. Rumelhart, David E, Hinton, et al. Learning representations by back-propagating errors[J]. 1986, 323(6088):399-421.

11. White H. Economic prediction using neural networks: the case of IBM daily stock returns[J]. IEEE International Conference on Neural Networks,1988.

12. Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

13. Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.

14. Yu Kai, Jia Lei, Chen Yuqiang, Past, today and future of deep learning[J]. Research and development of computer, 2013, 50(9): 1799-1804.

15. Chen Kai, Zhou Yi, Dai Fangyan. A LSTM-based method for stock returns prediction: A case study of China stock market[C]. IEEE International Conference on Big Data. IEEE, 2015:2823-2824.

16. Ding Xiao, Zhang Yue, Liu Ting, et al. Deep learning for event-driven stock prediction[C]// International Conference on Artificial Intelligence. AAAI Press, 2015:2327-2333.

17. Jia Hengjian. Investigation into the effectiveness of long short term memory networks for stock price prediction[J]. arXiv preprint arXiv:1603.07893, 2016.

18. Nelson D M Q, Pereira A C M, de Oliveira R A. Stock market's price movement prediction with LSTM neural networks[C]//Neural Networks (IJCNN), 2017 International Joint Conference on. IEEE, 2017: 1419-1426.

19. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

20. Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

21. Breiman L. Bagging predictors[J]. Machine learning, 1996, 24(2): 123-140.

22. Efron B, Tibshirani R J. An introduction to the bootstrap[M]. CRC press, 1994.

23. Hansen L K, Salamon P. Neural network ensembles[J]. IEEE transactions on pattern analysis and machine intelligence, 1990, 12(10): 993-10