Global Scientific JOURNALS

GSJ: Volume 9, Issue 3, March 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

Financial Time Series Analysis Using Machine Learning

Dost Muhammad

CS&IT, UET, Peshawar, Pakistan
dost@uetpeshawar.edu.pk

ABSTRACT

Analysis of finical time series is a complex task and an active area for researchers and financial analysts. Machine learning approaches have been used for financial time series analysis in recent years. Before the technological revolution, the researcher and experts of financial time series were used different traditional methods for prediction and forecasting purposes. Few years ago, the researchers applied several machine learning techniques on different data for analysis of financial time series. In this paper, we have discussed in detail multiple advanced machine learning algorithms and their model's architecture for financial time series analysis. After the observation, the performance of our presented methods i.e. logistic regression (LR), Random Forest (RF), Support vector machine (SVM), neural network (NN), and deep neural network (DNN) are too much better than traditional approaches.

Key Words

Financial Time Serie, Time Series Prediction and Forecasting, Machine learning, Deep Neural Network, Random forest, Logistic Regression, Support vector Machine

Introduction

The sequence of numerical data points in time order is called time series[1], time may be daily, monthly, yearly etc. There are two types of time series, univariate and multivariate time series. In time series analysis, involves a single variable (scalar) is univariate time series and multivariate time series contains two or more variables. The analysis of times series applies to different areas such as natural sciences, economics sector, biological sciences and numerical sciences. Figure 1.1 presents simple time series.

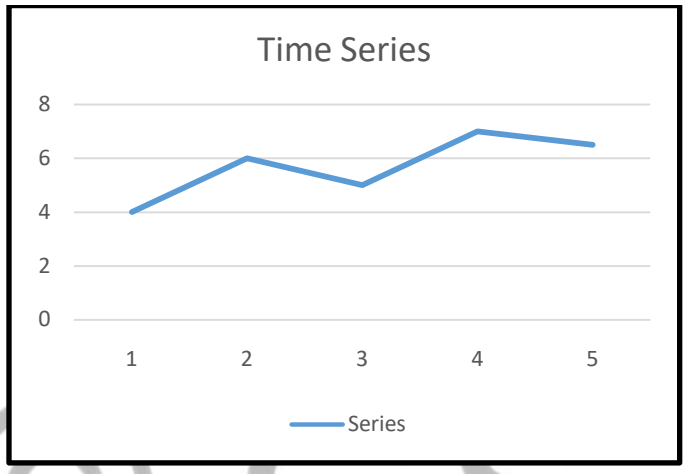

Figure 1.1 Time Series

The experts of natural sciences used several examples of time series i.e., the temperature of air, the level of water in a lake, the yields of corn in lowa and the size of natural population. In the economic sector, price index of consumers, unemployment ratio, gross national product, and population of the concerned area are used in the form of time series. The physical scientist used time series data for improvement the performance of radar, design and development of different rockets and airplanes. The mathematical theory of time series is the most interest area in chemical process and their investigations.

Nowadays, the analysis of financial time series data is the keen interest area for academia's and researchers. For time series analysis the experts were used multiple statistical models in the past years [2-5]. After the advent of advance machine learning approaches the stack holders of financial system are demanding for using the mentioned methods in stock markets analysis.

Machine learning is sub field of Artificial intelligence that learn from the past data. Machine learning has been used in a variety of applications such as fake news identification, clickbait detection, agriculture, health etc [6-8]. There are two types of

machine learning, supervised and unsupervised learning. The supervised learning mostly used for classification problems such as binary or multivariate classification. In this learning, the models were trained from dataset which contains historical data to classify positive class in the given data. The second type is unsupervised learning, the data is unlabeled and finding only patterns in data. In this studies, we have focused only on supervised learning techniques. We hope that our proposed approaches will be very helpful for the researcher and academia working on financial time series data.

Machine Learning Algorithms

Recall that machine learning approaches in financial time series analysis are powerful and very popular. The financial time series data are used for prediction and forecasting purposes. For aforementioned purposes, the performance of machine learning approaches are outstanding due to their complexity and advance learning nature. Multiple machine learning approaches are discussed in the following sections.

1) Logistic regression

The machine learning technique Logistic regression is mostly used for the analysis of binary and multivariate data. Logistic regression are basically using for understanding the data collected from multiple discipline such as medical and health, social sciences, economics [9] etc. The researchers used logistic regression when variables (dependent) are binary or categorical for prediction and classification purpose[9]. Logistic regression mathematically defined as follow.

Given data,

$$y = \{0,1\}, \qquad and\ we\ want\ 0 \leq h_\theta \leq 1$$

$$h\theta(x) = \theta^T x \qquad (1)$$

Let us modify Equation (1),

$$h\theta(x) = g(\theta^T x) \qquad (2)$$

After that, defining the logistic function also called sigmoid function in Equation (3)

$$g^{(z)} = \frac{1}{1 + e^{-z}} \qquad (3)$$

If we take Equation (1) and (2), and put it together so finally,

$$h\theta\ (x) = \frac{1}{1 + e^{-\theta^T x}} \qquad (4)$$

For two label classification, the performance of logistic regression is outstanding. In multi label classification "multinomial" extension and "cross-entropy" as a loss function are using for better performance.

2) Random forest (RF)

In supervised machine learning a flexible and easy to use algorithm is, random forest classifier. The role of random forest in financial time series analysis is well-known. The random forest is basically using the decision trees concept. The random forest can be used for regression and classification problems.

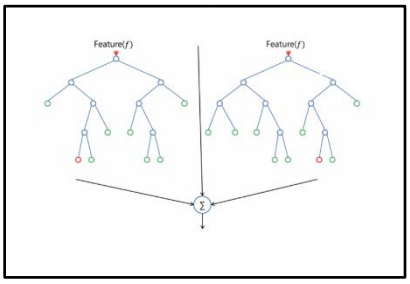

Figure 1.2 Random Forest

Random forest have no need for tuning the hyper parameter, mostly giving the great results. Random forest used their multiple trees for classifying the positive class and predict the actual labels. Using random forest for binary classification, GINI index is using as cost function, mathematically is following[10].

$$G_{index} = 1 - \sum_{i=1}^{c} (P_i)^2$$

To classify the actual class for multi class, the random forest used entropy as cost function which is defined in the following equation.

$$E_{nt} = \sum_{i=1}^{c} -filog(fi)$$

To making the model powerful then increase the number of trees in random forest classier, for good predictive purpose. "N job" and random state can be used increasing the speed of the random forest model.

3) Support Vector Machine (SVM)

A machine learning based algorithm which using in analysis of financial time series is support vector machine. Support vector machine addressing classification and regression problem, but mostly SVM is using for classification challenges. The objective of the support vector machine is to draw the best line

between the two or more than two classes. The created line separates the multi class labels from each other in the classifier. The following equation shows the cost function for support vector machine classifier[11].

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n} \theta_j^2 \#$$

Then;

$$\theta^T x^{(i)} \geq 1 \quad if \quad y^{(i)} = 1$$

$$\theta^T x^{(i)} \leq -1 if \quad y^{(i)} = 0$$

The support vector machine is furthermore divided into two types such as linear SVM and non-linear SVM. The linear support vector machine is mostly used for binary or two labels issues i.e. regression and classification. The linear SVM separates the data linearly and draw the line between dataset. Afterwards, when the data are not classified by using the single line and the data is arranged in the dataset non-linearly then the non-linear SVM will be performed better. In support vector model, for binary classification the decision function will be use "C=1" and for multivariate classification the "ovo" will be used as mentioned function.

4) Neural Network

After the outstanding performance in the engineering and cognitive sciences, the neural network plays an important role in financial time series analysis. The neural network is a powerful technique which is inspired from the human brain, using for classification as well as regression problems.

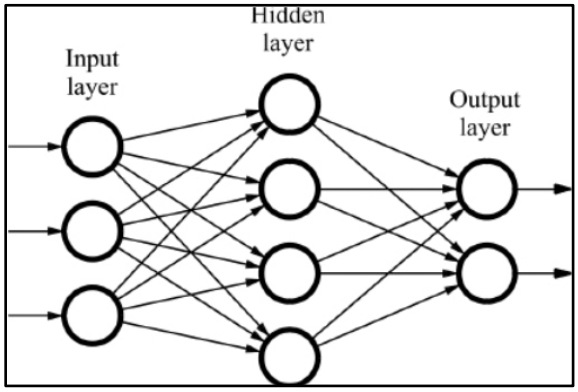

Figure 1.3 Neural Network

The neural network consist of input layer, hidden layer, output layer, activation function weights, bias and nodes. In hidden layer various nodes are used and it refer

the data to activation function. After the implementation of the activation function the model calculate the output. The neural network used only one hidden layer in the model for analysis, prediction and forecasting purposed. The mathematical concept behind the neural network is following [12].

$$Nn = f(b + \sum_{i=1}^{n} x_i\, w_i)$$

The above equation represents input (x_1 to x_n), their weights (w_1 to w_n), bias (b) and the activation function applied to the weighted sum of the inputs.

The parameters of the neural network model for binary and multi label classification is different from each other. For multivariate classification the activation function "softmax" and loss function "categorical_crossentropy" will be used. To predict the positive class, for binary classification "sigmoid" function will be used as activation and "binary_crosentropy" will be used as loss function in the model.

5) Deep Neural Network

A machine learning approach, deep neural network (DNNs) applied supervised and unsupervised learning problems. The deep neural network is the advance architecture of classical neural network. The performance of deep neural network in the field of financial time series are tremendous. The researchers[13, 14], applied the mentioned technique on multiple financial time series data and achieved outstanding results.

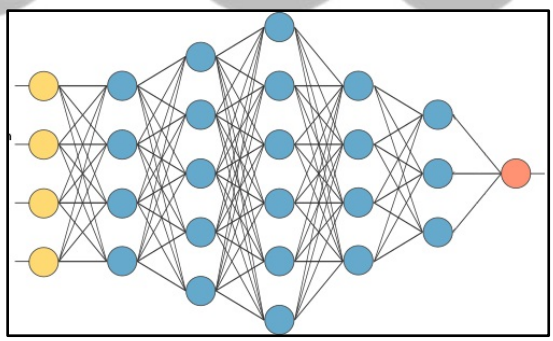

Figure 1.4 Deep Neural Network

The figure 1.4 represents there is more than on hidden layers in the model, which make the model complex and achieved desired results with good accuracy. Deep neural network are shown their performance in the finance sector in recent years and the experts declared the aforementioned approach for classification and regression issues. The main difference between classical neural network and deep neural network is only many hidden layers. The parameters which we have discussed in neural network section can be used for deep neural network such as activation and loss functions.

Conclusion

As we have discussed, the performance of machine learning approaches in field of financial time series prediction and forecasting is outstanding. In this study, we presented different methods which are using or best fitting for financial time series data such as LR, RF, SVM, NNs and DNNs. The methods RF, SVM, NNs and DNNs are suited for both classification and regression problem. Furthermore, the performance of logistic regression in binary classification is satisfactory. For multi label classification the performance of random forest and deep neural network are prominent. The performance of linear support vector machine in two label classification is remarkable

References:

[1] W. A. Fuller, *Introduction to statistical time series*. John Wiley & Sons, 2009.

[2] W. W. Wei, "Time series analysis," in *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*, 2006.

[3] M. P. Harrigan *et al.*, "MSMBuilder: statistical models for biomolecular dynamics," *Biophysical journal*, vol. 112, no. 1, pp. 10-15, 2017.

[4] J. V. Hansen, J. B. McDonald, and R. D. Nelson, "Time series prediction with Genetic-Algorithm designed neural networks: An empirical comparison with modern statistical models," *Computational Intelligence*, vol. 15, no. 3, pp. 171-184, 1999.

[5] R. S. Tsay, *Analysis of financial time series*. John wiley & sons, 2005.

[6] I. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, "Optimizing Pretrained Convolutional Neural Networks for Tomato Leaf Disease Detection," *Complexity*, vol. 2020, p. 8812019, 2020/09/23 2020, doi: 10.1155/2020/8812019.

[7] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, p. 8885861, 2020/10/17 2020, doi: 10.1155/2020/8885861.

[8] I. Ahmad, M.-A. Alqarni, A.-A. Almazroi, and A. Tariq, "Experimental Evaluation of Clickbait Detection Using Machine Learning Models," *Intelligent Automation \& Soft*

Computing, vol. 26, no. 6, pp. 1335--1344, 2020. [Online]. Available: http://www.techscience.com/iasc/v26n6/41030.

[9] S. Menard, *Applied logistic regression analysis*. Sage, 2002.

[10] M. Nelson, S. Kardia, R. Ferrell, and C. Sing, "A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation," *Genome research,* vol. 11, no. 3, pp. 458-470, 2001.

[11] V. Kecman, "Support vector machines–an introduction," in *Support vector machines: theory and applications*: Springer, 2005, pp. 1-47.

[12] T. Babs. "The Mathamtics of Nueral Network." Mediam. https://medium.com/coinmonks/the-mathematics-of-neural-network-60a112dd3e05 (accessed July 14, 2018).

[13] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," in *2016 7th International conference on cloud computing and big data (CCBD)*, 2016: IEEE, pp. 87-92.

[14] A. Navon and Y. Keller, "Financial time series prediction using deep learning," *arXiv preprint arXiv:1711.04174,* 2017.