Comparison of BPA-MLP and LSTM-RNN for Stocks Prediction

Roger Achkar, Fady Elias-Sleiman, Hasan Ezzidine, Nourhane Haidar
Department of Computer and Communications Engineering
American University of Science and Technology
Beirut, Lebanon
e-mail: rachkar@aust.edu.lb, fady_sleiman@hotmail.com, hasanezzidine@gmail.com, nourhanehhaidar@gmail.com

***Abstract*— Neural networks is considered one of the most developed concept in artificial intelligence, due to its ability to solve complex computational tasks, and its efficiency to find solutions. There is a wide range of applications that adopt this technique, one of which is in the financial investment issues. This paper presents an approach to predict stock market ratios using artificial neural networks. It considers two different techniques- BPA-MLP and LSTM-RNN- their potential, and their limitations. Tests were conducted on different data sets, such as FacebookTM stocks, GoogleTM stocks, and BitcoinTM stocks. We achieve a best-case accuracy of 97% for MLP algorithm, and 99.5% for LSTM algorithm. While the results appear to be promising, a web interface is presented in order to accept a certain amount of money, and accordingly checks the best stock to invest in.**

Keywords-Back propagation Algorithm, Multi-layer Perceptron, Long Short-Term Memory, Recurrent Neural Networks

I. INTRODUCTION

Neural networks have seen a huge interest in the past few years, with it being successfully applied in different areas such as finance, medicine, engineering, geology, and physics. Neural networks attempt to mimic the fault tolerance and learning capacity of the brain [1]. In the below, neural networks were adopted in price forecasting for stocks. Stocks has been a more accessible investment tool, not only to investors but also for common people. One characteristic that all stocks have in common is their uncertainty. The aim of the project is to reduce this uncertainty using neural networks. Neural networks are mathematical structures designed to mimic the architecture, fault-tolerance, and learning capability of the human brain [2]. The below two sections will briefly explain two adapted techniques in neural networks, MLP and LSTM in RNN, in the closing price stocks market. The tests are conducted on FacebookTM stocks, GoogleTM stocks, and BitcoinTM stocks. Required data were collected from Yahoo finance [3].

II. LITERATURE REVIEW

Researchers and specialists have demonstrated keen interest in studying the consistency of the stock market. This interest started to rise first in 1900 when Louis Bachelier, a French mathematician, was the first to talk about the random characteristics of share price behavior [4]. From that point forward, the number of researches trying to study the efficiency of stock market was on rise, with each study contributing in its own way to understand the behavior and the predictability of the markets. One of the recent studies that contributed in a significant way was the time-series forecasting. A model done by Fan and Yao in 2003was executed in a way that the forecasting was performed using statistical-based methods, for example ARMA (autoregressive moving average), because of its flexibility to model many stationary process [5]. This was later extended by Weron and Misiorek in 2008 [6]. However, the ARMA model assumes a linear relationship between the lagged variables, thus produces an approximation to real-world complex systems but fails to predict the evolution of nonlinear and nonstationary processes. Accordingly, ARIMA (autoregressive integrated moving average) was later used to remove/reduce first-order non-stationarity, while the ARCH model was used to capture the second-order non-stationarity. Even though these methods have gone far from where they began, they are still limited [7] [8]. That is why, over the past few decades' artificial neural networks have attracted tremendous attention in the time-series forecasting community. Compared to the statistics –based methods mentioned above, the neural networks approach showed some unique characteristics such as being both nonlinear and data driven, being nonparametric, and being more flexible and universal. In the below sections, two methods in neural networks will executed, the back-propagation algorithm in multilayer perceptron, and the long-short term memory in recurrent neural networks

III. MULTILAYER PERCEPTRONS

Multilayer perceptron, also called feed-forward networks, are the most popular type of neural networks in use today. In MLP structure, the neurons are called layers, with the first layer being the input, and the last is the output. The remaining layers are called hidden layers.

In a neural network, each neuron- with the exception of the input neuron- received and processes inputs from other neurons. This is called information processing by a neuron, where the information is thus presented at the output of the neuron, as demonstrated in Fig 1.

978-1-5386-9450-3/18/$31.00 ©2018 IEEE
DOI 10.1109/ISCBI.2018.00019

A. Activation Function

In the methodology proposed, the activation function used is the hyper-tangent function between -1 and 1. This was preferred over logistic functions due to the fact that it is symmetric over 0, which makes the training easier and more accurate.

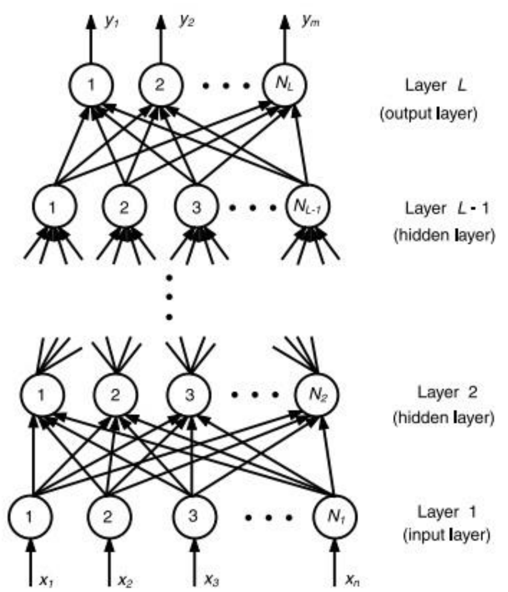

Figure 1: Multilayer Perceptron Structure [9]

B. Weights and Bias

The weights and bias are initially generated randomly. With every back propagation, the weights are corrected accordingly

C. Number of Neurons and Layers

MLP is divided into three main layers: input, output and hidden layers. However, there is no specification on the size of the network (i.e. number of hidden neurons). Average hidden neurons should be concluded depending on the degree on nonlinearity and the dimensionality of the problem stated. Too much hidden neurons will lead to over learning of the neural network. On the other hand, too little neurons will not give the network the freedom to learn accurately. Moreover, it is necessary to have at least one hidden layer, but it is sufficient to have more than one. One hidden layer may require too many hidden neurons, so in practice, two hidden layers would perform better in modeling a nonlinear problem [10] [11] [12] [13] [14]. After several experiences, the paper will follow a topology of 7-3-2-1 neurons.

The network is starting with seven inputs. Seven days will be considered into order to predict the eighth.

D. Training Process

The approach done using this algorithm is to learn from past results in order to predict the future. The first step is to initialize a learning rate of 0.02. One of the examples experimented is Facebook^{TM} stocks. Stock prices from 2012 were extracted and tested on. The first step is to cross validate this data, and divide them into two parts, a large part for training, and another small part for testing. Afterwards, the first seven data were taken into consideration as inputs. The system thus predicts the eighth, and compare it to the actual data we have on the eighth day to correct the weights accordingly. Then, the seven data inputs will slide so that the system starts from day two till day eight, in order to predict the ninth. This process will be repeated until we reach the last training input day, were each propagation provides a better learning experience. Everything afterwards till the present day will be for testing purposes and to check accuracy.

Figure 2 shows the training process of the Facebook stocks in BPA. the blue line shows the real data, the Orange line shows the trained data, while the yellow line shows the predicted data. By comparing the predicted data to the real data, it is noticeable that they are very close and the system is valid with a small percentage error.

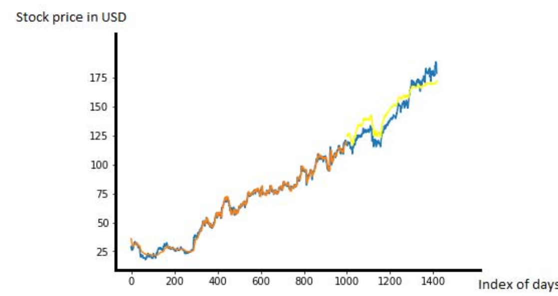

Figure 2. Facebook Training in BPA

IV. LSTM – RNN

The following approach is the long short term memory in recurrent neural networks. This approach is considered due to the fact that it considers reasoning, unlike the MLP. After testing, a comparison will be held between both methods to determine the one with better performance.

In order to explain the RNN, imagine you want to classify what kind of event is happening at every point in a movie. It's unclear how a traditional neural network could use its reasoning about previous events in the film to inform later ones. Recurrent neural networks address this issue. Different from the feed-forward neural networks represented by MLP, RNN allows the connection of the network to form a cycle, allowing information to persist.

On the other hand, LSTM networks are special kind of RNN that are capable of learning long-term dependencies. These networks introduce the memory cell that replaces the traditional artificial neurons in the hidden layers of the network. With memory cells, networks are able to effectively associate memories, hence grasp the structure of data dynamically over time with high prediction capacity. The basic unit of the LSTM architecture is a block memory with one or more different types of memory cells and three adaptive multiplications called input gate, forget gate and output gate.

The input and the output gates multiply the input and output of the cell while the forget gate multiplies the cell's previous state. This is demonstrated in Fig. 3. The activation function of the gates is usually the logistic sigmoid function so that it is between 0 (gate closed) and 1 (gate open). The activation function of the input or output of the cell is usually tanh or logistic sigmoid.

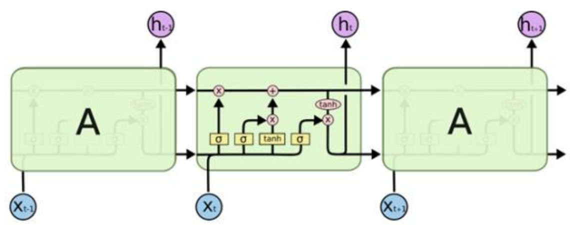

Figure 3. LSTM memory block with a single cell

A. Adam Algorithm

Adam stands for Adaptive Moment Estimation. Adam can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data. Some of the characteristics of the algorithm is the fact that it is:

- Straight forward to implement.
- Computationally efficient.
- Little memory requirements.
- Invariant to diagonal rescale of the gradients.
- Well suited for problems that are large in terms of data and/or parameters.
- Appropriate for non-stationary objectives.
- Appropriate for problems with very noisy/or sparse gradients.
- Hyper-parameters have intuitive interpretation and typically require little tuning.

Adam was demonstrated empirically to show that convergence meets the expectations of the theoretical analysis. Adam was applied to the logistic regression algorithm on the multilayer neural network, to conclude that using large models and datasets, we demonstrate Adam can efficiently solve practical deep learning problems.

B. Keras Library

Keras is a high-level neural networks application programming interface developed with a focus on enabling fast experimentation. It allows for easy and fast prototyping and supports recurrent networks. The LSTM tests are implemented using this library, where the system is based on four layers, while the neuron numbers, the learning rate, the weights, and the tuning momentum are automatically generated based on the problem issued.

After presenting the data into the program developed using Keras library, the results in Fig. 4 were obtained. The blue line represents the data provided, the orange line represents the trained data, while the yellow line shows the predicted data. Comparing the yellow line to the blue one, it is notable that the prediction is valid with a small error.

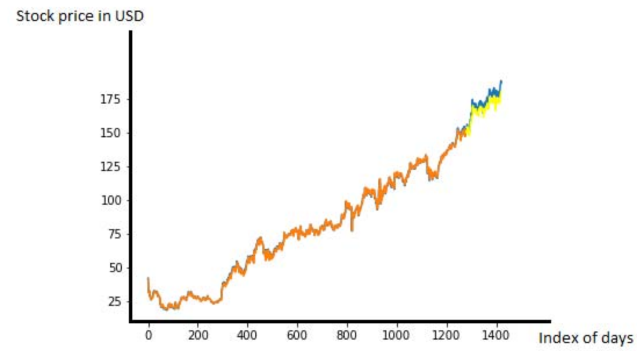

Figure 4: Facebook training in LSTM

V. Results

Post learning and testing, we achieve the results shown in Figures 5 and 6. The predicted error value in BPA fluctuated between 3% and 16% as when calculating the round mean error squared, while in LSTM it fluctuated between 0.5% and 12%. The results are close to each other, however we notice LSTM achieves a better result.

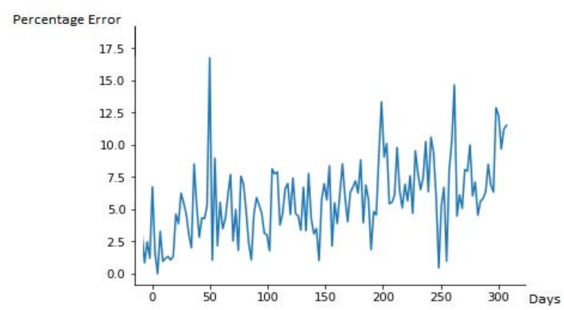

Figure 5. Percentage Error in BPA

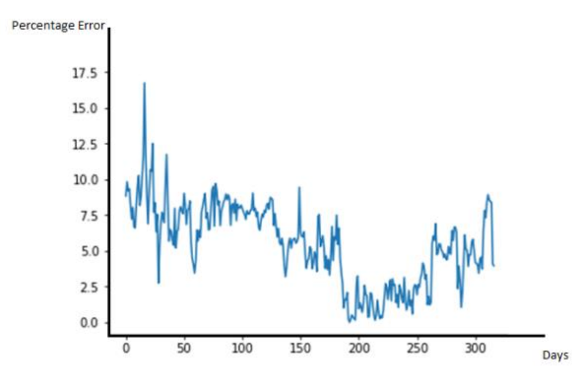

Figure 6. Percentage Error in LSTM

VI. Demonstration

The best way to put the above studies into practice, was to show a real-life demonstration. Both algorithms were tested by a program code done using Python and JavaScript.

It is clear that the LSTM network is a better approach in neural networks than that of the BPA, since the percentage error was slightly less. In this case, we formed a calendar using HTML SCC to observe the outcome when investing a certain amount of money into stocks, using the LSTM approach.

Figs. 7, 8 and 9 show respectively the investing of Google, Facebook and bitcoin in the months of February, March and April. Although we are only using the data of these three stocks, the program can be altered to have any stock data analyzed. The interface intends to show you the stocks that will get you the highest profit on daily basis, and can calculate for you the estimated monthly profit of a certain stock given a certain amount invested. For example, Fig. 7 is displaying the overall monthly profit when you invest with a 1,000 USD in Google stocks in February. It also displays the average daily profit and the cumulative asset. On the other hand, Fig. 8 shows the profit when investing in 1,000 USD in Facebook stocks in March.

Such analysis can be repeated for different stocks, and for more time-periods. The program developed is able to show the forecast of stocks for any time of the future, while also showing the cumulative asset given a specific time interval whether it is only one month or more.

VII. Conclusion

In this paper, two different approaches in neural networks were performed in order to predict the stocks rate. We described the theory behind back propagation algorithm and recurrent neural networks, to be able to construct a stable program that could learn from past data, the future of given stocks. Both used techniques were successful and fairly accurate.

As is evident from Figs. 5 and 6, the percentage error in LSTM is slightly less than that in BPA. Even though BPA is much simpler to assemble, an investor would definitely yield to the method with less percentage error.

To conclude, it is believed that neural networks prove to be an effective and promising tool for stock market prediction.

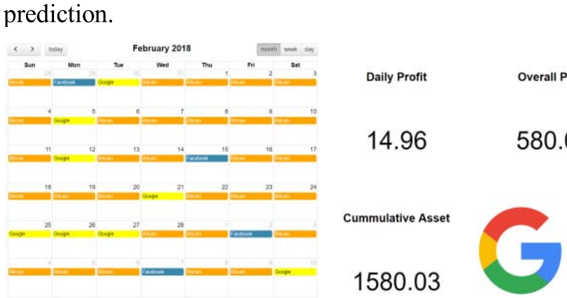

Figure 7. Investing in Google in February

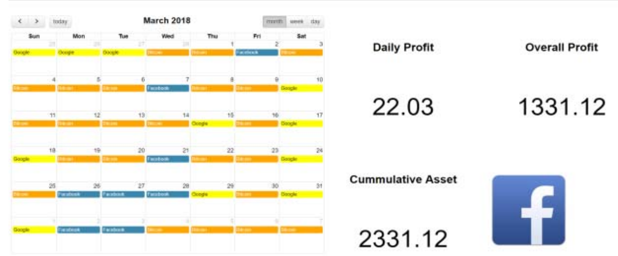

Figure 8. Investing in Facebook in March

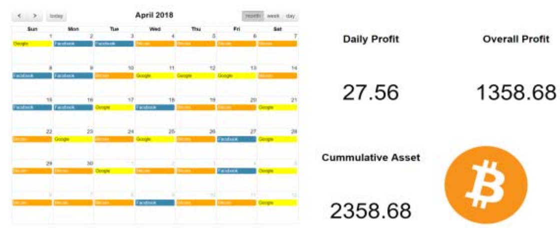

Figure 9. Investing in Bitcoin in April

References

[1] A. Kar, in *Stock Prediction using Artificial Neural Networks".*

[2] A. Munasinghe and D. Vlajic, in *Stock market prediction using artificial neural*.

[3] Y. Website, "https://uk.finance.yahoo.com/," [Online].

[4] D. M. H., "Louis Bachelier's "Theory of Speculation"," Imperial College.

[5] F. J. a. Y. Q., "Nonlinear Time Series: Nonparametric and Parametric," New York, NY., 2003.

[6] W. R. a. M. A., "Forecasting spot electricity prices: a comparison of," International Journal of Forecasting, 2008.

[7] E. R.F., "Autoregressive conditional heteroscedasticity with estimates of the," Econometrica.

[8] T. H., "Non-Linear Time Series: A Dynamical System Approach," Oxford.

[9] K. C. Gupta, Q. J. Zhang and K. C. Gupta, in *Neural Networks for RF and Microwave Design*, 2000.

[10] C. Saide, R. Lengelle, P. Honeine, C. Richar and R. Achkar, "Nonlinear adaptive filtering using kernel-based algorithms with dictionary adaptation. International Journal of Adaptive Control and Signal Processing," 2015.

[11] G. A. Kassam and R. Achkar, "LPR CNN Cascade and Adaptive Deskewing" Complex Adaptive Systems, Procedia Computer Science, Elsevier, Volume 114, November 2017, pp 296-303.

[12] M. Owayjan, R. Achkar and M. Iskandar, "Face Detection with Expression Recognition using Artificial Neural Networks," Beirut, Lebanon, 3rd Middle East Conference on Biomedical Engineering (MECBME), October 2016, pp. 116-120.

[13] R. Achkar, M. ElHalabi, E. Bassil, R. Fakhro and M. Khalil, "Voice identity finder using the back propagation algorithm of an artificial neural network," 2016.

[14] Y. Harkouss, S. Mcheik and R. Achkar., "Accurate wavelet neural network for efficient controlling of an active magnetic bearing system," 2010.