Stock Market Cost Forecasting by Recurrent Neural Network on Long Short-Term Memory Model

Srikanth Tammina

Dept. of Electrical Engineering, Indian Institute of technology, Hyderabad – Telangana, 502285

Abstract - *Thoughts of human beings are determined; they don't think from scratch for every problem they encounter. We read this paper and comprehend each sentence and paragraph based on our understanding of previously learned words. We don't dump everything and start thinking from scratch. Conventional feed forward vanilla neural network cannot remember the thing it learns. Each iteration starts fresh during training and it doesn't remember the instances in the previous iterations. The shortcoming of this kind of learning especially while dealing with time series data is it avoids any correlations and data patterns in the dataset. Recurrent neural network is a kind of artificial neural network that recognizes the correlations of past iterations and predict the next sequence of actions. This paper takes the advantage of one of the most successful Recurrent neural network architectures: Long Short-Term Memory to predict instances in future.*

Key Words: Recurrent neural network, Long Short-term memory, time series, and machine learning.

1. INTRODUCTION

In various domains like manufacturing, telecom – analyzing and predicting of time series data has always been an imperative technique in a wide list of problems, including demand forecasting, sales forecasting, road traffic administration, economic forecasting and earthquake forecasting. Conventional neural network cannot handle the above-mentioned time dependent events. To counter this problem a new powerful model is built from recurrent neural network – Long Short-Term memory. LSTM introduces the memory cell, a unit of computation that substitutes conventional artificial neurons in the hidden layer of the network. With these memory cells, networks can effectively associate memories and input remote in time, hence, suit to grasp the structure of data dynamically over time with high prediction capacity [14].

Stock market prediction is the act of guessing to govern the future value of a business or a company stock or other financial instrument traded on a financial exchange. The effective and near accurate prediction of a stock's future price will maximize investor's profits. Study to predict stock market has been undertaken by many wide range research groups, aiming at improving the accuracy and counter the challenges in prediction.

The scope of stock market prediction varies in three ways [1]:

I. The targeting price change can be near-term (less than a minute), short-term (tomorrow to a few days later), and long-term (months later)

II. The set of stocks can be limited to less than 10 particular stock, to stocks in a particular industry, to generally all stocks.

III. The predictors used can range from global news and economy trend, to particular characteristics of the company, to purely time series data of stock price.

Artificial neural networks are series of artificial neurons attached together as a network where each neuron performs a single computational job and combinedly creates an intricate model which can learn to recognize patterns in the dataset, extract relevant features and predict the outcome. On a contrary, deep neural network are a sort of neural networks with more than one hidden layer of neurons. Based on the number of the layers in the deep neural network, the embedded layers allow it to represent data with a high level of abstraction [11]. Deep neural network still faces shortcomings while trying to predict time series data such as demand forecasting, stock market, traffic management because these networks are still inadequate in their capability to predict time dependent data. When working with time dependent data or time series data, deep neural network cannot process dependent time series at every step and cannot remember the entire previous time steps. To counter the problem, a new successful neural network - Recurrent neural networks (RNN) were proposed by John Hopfield in 1986. Recurrent neural network deal with one time step at a time and allows the input information to the neural network across multiple time steps. RNNs preserve the input information hence, resolving the problem of overlooking previous input information. First generation recurrent neural network performs efficiently in modelling short time sequences. These neural networks have hidden layers which take the current iteration of input and consider the previous iteration's hidden layer output as inputs. The downside of modelling in this way is to encounter a new problem of vanishing gradient. Recurrent neural network suffers from the problem of vanishing gradient which inhibits learning of long data sequences. Vanishing gradient occurs when memory of older data becomes subdued, causing the gradient signal to become very small hence, model learning can become very slow for long term time dependencies in the data [7]. Contrastingly, if the weights in the hidden layer matrix becomes large, this will eventually lead to a scenario where the gradient is so large that learning scheme diverges from the ground truth. To tackle this problem of vanishing gradient, a novel approach called Long Short-Term Memory (LSTM) model was proposed by Sepp Hochreiter and Jürgen Schmidhuber in 1997. Contrast to vanilla recurrent neural network, Long Short-Term Memory network substitutes neurons with a memory unit. The architecture inside a single step cell of Long Short-Term Memory constitutes a hidden state, an input and an output. Inside this single step cell, there are three gating units called: the input gate, the forget gate and the output gate. Input gate and output gate perform the same role as simple vanilla recurrent neural network. Based on the input data, the forget gate learns to memorize weights of previous time instances [7]. The vanishing gradient issue can be averted by feeding the internal state across multiple time steps with weight of 1 hence, this precludes the vanishing gradient problem since any error that passes into the recurrent network unit during learning stage is stored indeterminately. Input gate decides when to let activation pass into the cell, each gate has a sigmoid activation function. Output gate learns when to let activation pass out of it to the rest of network. During back propagation, error flow into the cell will be decided by output gate and letting error flow out of the cell will be decided by input gate.

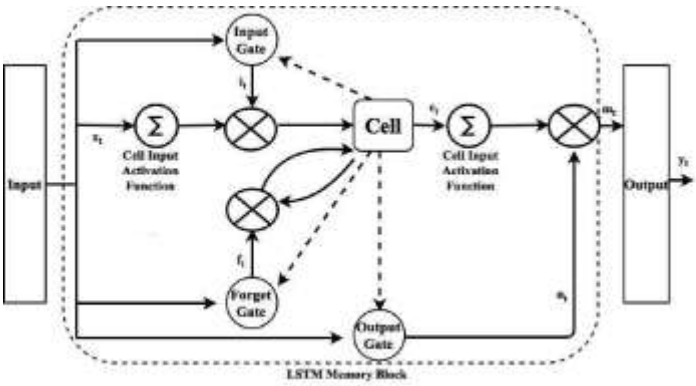

Fig. 1. Architecture of Long Short-Term Memory [5]

Thus, Long Short-Term memory network are preferably employed in capturing how variation in one particular stock price can gradually affect the cost of various stocks over longer periods in time.

1.1 Advantages Of LSTM

- LSTM in general a great tool for anything that has sequence. Since the meaning of the word depends on the ones that preceded it.
- The constant error propagation within memory cells results in LSTM ability to bridge very long time lags.
- For long time lag problems, LSTM can handle noise efficiently. In contrast to finite state automata or hidden Markov models LSTM does not require a prior choice of a finite number of states. It can deal with unlimited state numbers.
- LSTM does not require any parameter tuning. LSTM work well with wide range of parameters such as learning rate, input gate bias and output gate bias.
- LSTMs are used in generating text in text analytics. Once the LSTM model is trained on the corpus of a writer, the model will be able to generate new sentences that mimics the style and interests of the writer

2. DATASET

Dataset contains historical stock prices (last 5 years) for all companies currently found on the S&P 500 index. All files have the following columns:

- Date – data of the stock price recorded
- Open - price of the stock at market open (this is NYSE data so all in USD)
- High - Highest price reached in the day
- Low Close - Lowest price reached in the day
- Volume - Number of shares traded
- Name - the stock's ticker name

	date	open	high	low	close	volume	Name
0	2013-02-08	15.07	15.12	14.63	14.75	8407500	AAL
1	2013-02-11	14.89	15.01	14.26	14.46	8882000	AAL
2	2013-02-12	14.45	14.51	14.10	14.27	8126000	AAL
3	2013-02-13	14.30	14.94	14.25	14.66	10259500	AAL
4	2013-02-14	14.94	14.96	13.16	13.99	31879900	AAL

Fig. 2. Table shows first 5 rows of data in training dataset

3. Leveraging LSTMs to predict stock price

In Recurrent neural network, while input is sent for processing to hidden layer then the words or input text gets transformed into machine readable vectors which will be processed by recurrent neural networks one at a time. During processing these vectors, RNN passes the previous hidden state to the next step of the sequence. This hidden state act as a memorized block which holds the information of the previous data which network has processed before. The current input and previous hidden state combine and passes into tanh activation function. Tanh activation function is employed to help normalize the values passing through the network.

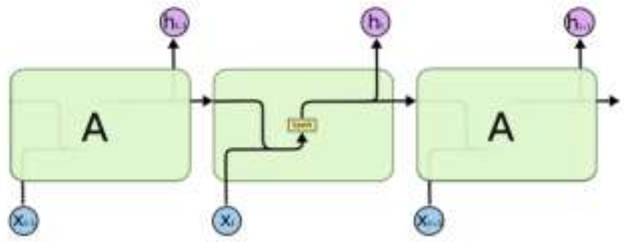

Fig. 3. Recurrent neural network single cell architecture

In the above image, Xt is the current input, ht-1 is the previous hidden state, Xt-1 is previous input. In current state, the input to the tanh function is current state input Xt and previous hidden state ht-1. The input and hidden state combined to form a vector and information from this vector passes through tanh function and output from the tanh function is new hidden state. This new hidden state passed as an input hidden state to Xt+1.

$$h_t = f_W(h_{t-1}, x_t)$$

$$\downarrow$$

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

Fig. 3. Simple vanilla RNN's present state cell

ht = hidden state which is a function of previous hidden state ht-1

yt = Element wise multiplication of Why and ht

Recurrent neural networks are successful and efficient in handling sequential data, but there are limitations for RNN while dealing with long term dependencies in the input data. **For example: I live in India and I speak in ___.** The answer must be **'Hindi'** here but if there are some more words in between '**I live in India & 'I speak in ___**'. It'll be difficult for recurrent neural networks to predict 'Hindi. This is the problem of Long-Term Dependencies. Hence, we leverage the use of LSTMs to counter these kinds of problems.

Memorizing information for longer periods of time is the core functionality of these networks. Long Short-Term Memory network are similar to RNNs, but the repeating module has completely different structure compared to simple vanilla RNN cell.

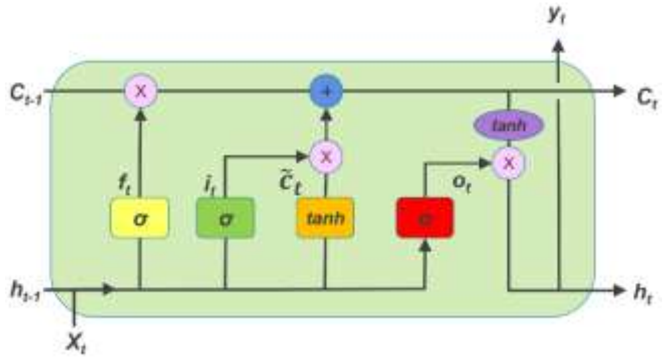

Fig. 4. Single cell LSTM architechture

- **Forget gate –**

$$f_t = \sigma_g(W_f x_t + U_f c_{t-1} + b_f)$$

Once the input (previous state ht-1) is passed into forget stage, forget gate decides which error needs to exclude from ht-1 state and preserving related information. Forget gate constitutes of a sigmoid function which throws an output in ranges between (0,1).

- **Input gate –**

$$i_t = \sigma_g(W_i x_t + U_i c_{t-1} + b_i)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + b_c)$$

New information will be passed into input gate and decides which information in input data needs to store. Sigmoid function (σg) decides which values need to be updated and tanh layer creates a vector of new candidates (ct) which needed to be included in present cell state.

- **Output gate –**

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

This gate decides what information needs to output from previous cell state. Input will be multiplied with tanh function to get the output in range between (-1,1). The output of this will be multiplied with sigmoid function.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 09 | Sep 2019 www.irjet.net p-ISSN: 2395-0072

4. Algorithm Pseudo Code

- **STEP 1**: Initialize the problem by importing data set – Historical stock price data of S&P 500 is imported
- **STEP 2**: Prepare training and test data - Split the data into training data which contains 80% of the total dataset and 20% as test data
- **STEP 3**: Model building –
    - Include LSTM function with input shape (7,1)
    - Add dense layer of 1
    - Reshape the model for train and test
    - Fit LSTM model with historical data to check for overfitting
- **STEP 4**: Run the model for 300 epochs and visualise the results
- **STEP 5**: Predict the model on test data
- **STEP 6**: Evaluate accuracy metrics by comparison on predicted data and test data

5. Results

5.1 Exploratory data analysis - Below are the univariate and multivariate analysis for several stock company prices

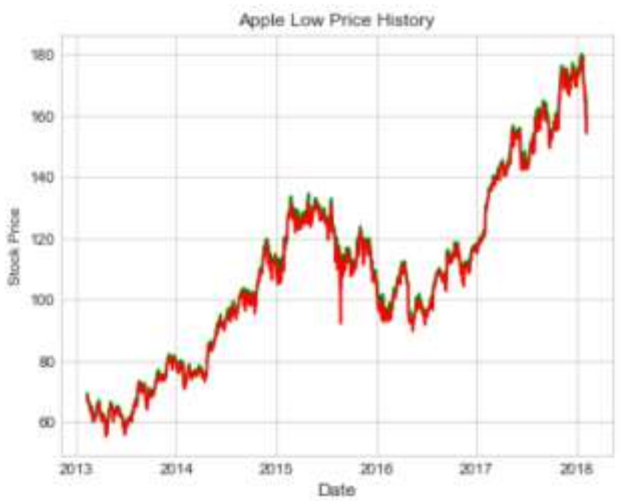

Fig. 5. Apple low price historical graph

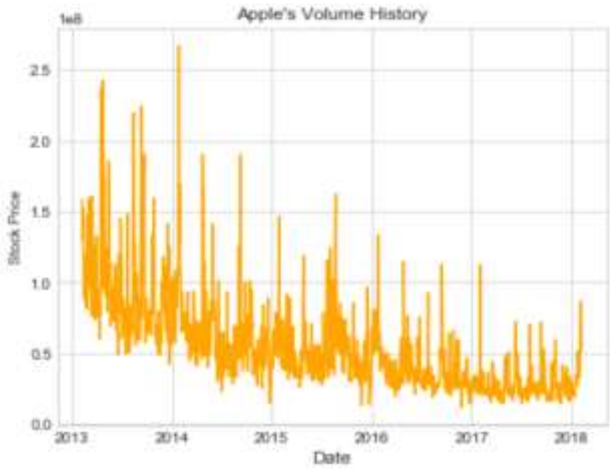

Fig. 6. Apple volume historical graph for periodical year between 2013-2018

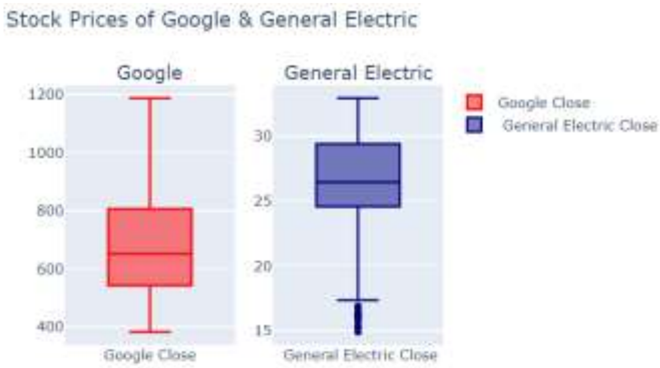

Fig. 7. Box plot to analyze the data patterns and outlies in Google and General electric prices

Fig. 8. Candle stick plot feauturing open, close, high and low prices for Google and General electric companies

5.2 Model Results

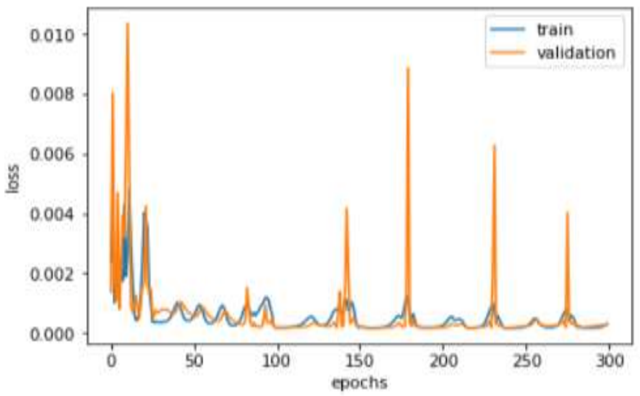

Fig. 9. Shows trainng and validation loss after running the LSTM model on training data and testing on validation data. Before training the model, input data is scaled between (0,1).

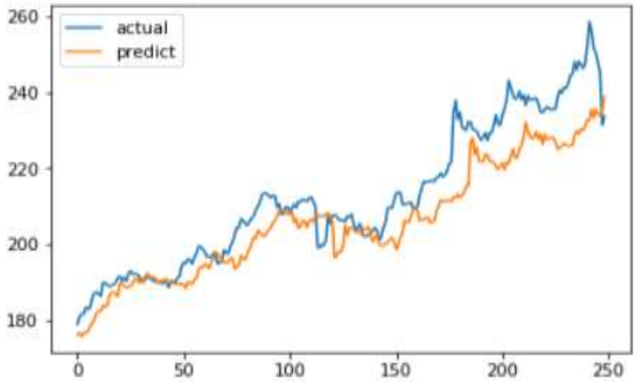

Fig. 10. Actual and Predicted closing stock price for **AbbVie Incorporation.** LSTM model trained on the batch size of 256 and ran total of 300 epochs.

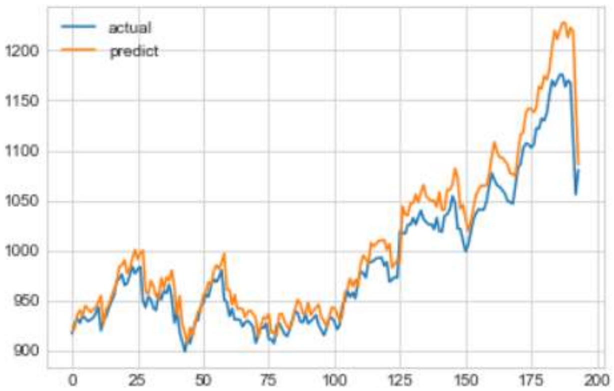

Fig. 11. Actual and Predicted closing stock price for **3M Company**. LSTM model trained on the batch size of 256 and ran total of 300 epochs.

```
predicted:[[918.7923]], actual:[[916.44]]
predicted:[[922.43604]], actual:[[927.04]]
predicted:[[935.1595]], actual:[[931.66]]
predicted:[[940.09595]], actual:[[927.13]]
```

Fig. 12. This image shows few data points of closing stock price prediction of **3M company** by testing it on LSTM model

6. CONCLUSION

Forecasting on stock market costs are a cumbersome and challenging task due to the fact that it is an enormously intricate and sophisticated. Numerous researches from disparate areas are targeting to tackle the challenge of predicting accurate future stock price inorder to cash out profits which is a primary objective for every stock trader. Because stock prediction depends on sequence prediction, LSTMs are prooven extremely effective. As LSTMs are able to store previous data which is imperative in this scenario and forget the irrelevant data.

Stock Market still depends on human intuition and emotions. More robust and accurate prediction includes new feed analysis from wide range of online media platforms like Reddit, Facebook and Twitter where sentiments can be extracted. In 2013, a scientific research paper published by Helen Susannah Moat and Tobias Preis in which they exhibited a correlation between changes in the number of views of English Wikipedia articles relating to financial topics and subsequent large stock market moves [7][8]. The use of text mining algorithms combine with deep learning techniques has receiver prodigious attention since last decade. By using Text sentiment analysis fused with RNNs and LSTMs to get more accurate predictions and decrease error.

REFERENCES

[1] Alice Zheng and jack Jin, Using AI to Make Predictions on Stock Market, 2018. [Pdf]

[2] E.F. Fama, B.G. Malkiel, "Efficient capital markets: A review of theory and empirical work", The Journal of Finance, vol. 25, no. 2, pp. 383-417, 1970.

[3] A.E. Biondo, A. Pluchino, A. Rapisarda, D. Helbing, "Are Random Trading Strategies More Successful than Technical Ones?", PLoS ONE, vol. 8, pp. e68344, Jul. 2013.

[4] B. Batres-Estrada, Deep learning for multivariate financial time series, 2015.

[5] Sucheta Chauhan, Lovekesh Vig. 2015 IEEE International conference on data science - Anomaly detection in ECG time signals via deep long short-term networks

[6] O.H. Luca, Di Persio, "Artificial neural networks approach to the forecast of stock market price movements", international Journal of Economics and Management Systems, vol. 1, pp. 158-162, 2016.

[7] K. Chen, Y. Zhou, F. Dai, "A LSTM-based method for stock returns prediction: A case study of china stock market", Big Data (Big Data) 2015 IEEE International Conference on, pp. 2823-2824, Oct 2015.

[8] Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y. Kenett, H. Eugene Stanley and Tobias Preis (2013). "Quantifying Wikipedia Usage Patterns Before Stock Market Moves". Scientific Reports. 3: 1801. doi:10.1038/srep01801. PMC 3647164.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[9] "Wikipedia's crystal ball". Financial Times. May 10, 2013. Retrieved August 10, 2013.

[10] Eugene F. Fama, "The Behavior of Stock Market Prices", the Journal of Business, vol. 2, no. 2, pp. 7-26, January 1965.

[11] S. V. Chang, K. S. Gan, C. K. On et al., "A review of stock market prediction with Artificial neural network (ANN)", IEEE International Conference on Control System Computing and Engineering. IEEE, pp. 477-482, 2014.

[12] A. M. Rather, A. Agarwal, V N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns", Expert Systems with Applications, vol. 42, no. 6, pp. 3234-3241, 2015.

[13] F. A. Gers, J. Schmidhuber, F. Cummins, "Learning to forget: continual prediction with LSTM", Neural Computation, vol. 12, no. 10, pp. 2451, 2000.

[14] K Greff, R K Srivastava, J Koutník, "LSTM: A search space odyssey[J]" in IEEE transactions on neural networks and learning systems, 2016.

BIOGRAPHIES

Srikanth Tammina graduated in Electrical Engineering from Indian Institute of Technology, Hyderabad. His research interests include Machine learning, Deep learning, Data mining and Internet of Things.