

Grab Your Yield on Stock Market Using Sentimental Analysis and LSTM using ARNN

B.S.Liya¹, S.Srividya², J.Roja³, V.Chandralekha⁴

Assistant professor, Department of CSE, Prathyusha Engineering College¹
Final year student, Department of CSE, Prathyusha Engineering College^{2,3,4}
liyabadhu13@gmail.com¹, srividyasree75@gmail.com², rojajules@gmail.com³,
lekhavetherath054@gmail.com⁴

Article Info
Volume 83
Page Number: 4709-4714
Publication Issue:
May-June 2020

Abstract

Stock market refers to an auction. Investors who can buy and sell shares of corporate stocks. The values of stock are very complex to predict. Moreover, accurate prediction of stock is greatly difficult task till now so that the stock price of a company depends on various parameters such as fame of the company and other sudden non-predictable events that may cause neural disasters etc. In our project we are predicting the stock with the help of Machine Learning and Artificial Recurrent Neural Network approach. In this we are having two approaches to predict the values of stock. First one is Advanced Long Short-Term Memory (LSTM) approach and second one Sentimental Analysis approach. Sentimental Analysis uses Twitter API to predict the values of stock. By above problem understanding, this product is developed to provide the accurate values of stock rather than giving the irrelevant stock values.

Article History
Article Received: 19 November 2019
Revised: 27 January 2020
Accepted: 24 February 2020
Publication: 16 May 2020

Index Terms: Long Short-Term Memory, Neural Networks, Twitter API, Sentimental Analysis.

1. Introduction

Stock market rate is one of the far most prominent fields in the sector of computer science, economics, accounting and management, etc. Stock market one day ahead has been a desire of many dealers. "To what scope can the history of a typical stocks cost be used to form valid predictions regarding the future price of the stock. However, information concerning a stock is normally uncertain, vague and incomplete making it is a dare to predict the future economic performance. Accurate stock market predictions are very important. Still after many attempts by scientists, No methods which can successfully predict the exact stock price can be created. The difficulty of prediction is due to the complexities of. modelling market dynamics per mining the stock market this is because stock price not only depends on the import/export of the company but also name and fame of the company. To predict stock value, we must plot the stock indexed with respect to time, this method of plotting time series on a graph is called as time-series forecasting.

This time series will be created as a model We train this model on the history of data which we have retrieved and use this model to predict the future values. This can be done with the help of Machine Learning algorithms. In this paper we are using the most popular and importantly used model Artificial Neural Network (ANN). ANN is the basic framework for most of the Machine Learning algorithms used till the date[1]. Among this framework we are using Artificial Recurrent Neural Network(ARNN). ARNN is the special kind of ANN which are powerful and robust type of network and belong to the most promising algorithm. Because of their internal memory, RNN have enough strong memory decide and remember the important contents from the input they acquires[4].LSTM is a most remarkable algorithm a time series based prediction due to its better remembering capabilities for Long and Short Term using their memory gates. In this research we also hypothesize that sentimental analysis is also used to give the effective results when certain non-

predictable even occurs by combining it with input variables(tweets). Moreover, sentimental analysis using live streaming of twitter data which cannot be done by the using LSTM approach to predict stock status.

2. Related Theory

Recurrent Neural Network

Recurrent neural network is a model where the output from previous unrolled batch has been fed as input to the current batch i.e., it fetches the previous store data and finds the unclear or noisy data which will be used in future points this process will not stopped until we reach our milestone[2][3].RNN structure consists of three layers input layer, hidden layer and output layer.

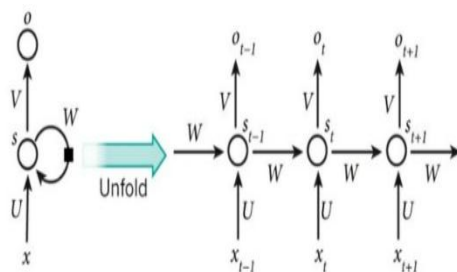


Figure 1: Batch unfolding

The input layer is the base layer which we used to train our machine for the prediction of undetermined noisy/uncleared data and hidden layer which is also known has noisy unclear data layer will be trained again to find for the data. This process can do recursively until we reach the theorem cleaner data which can be easily be learnt by machine [8]. RNN have a memory which collects and store information in short remembers all information about what has been calculated. Tis reduces the complexity of parameters, as the data is not clean and easily understandable by the machine unlike other neural networks.

The above diagram represent the unfolding of RNN network regarding time of the computation involved in its forward computation and the recurrent network behaviours can be described as:

$$s_t = fH(Ux_t + Ws_{t-1})$$

Where x_t is the input vector at time step t , s_t is the output at step t , V, W, U are weights, f, H is activation functions

Long-Short-Term-Memory

Long-Short-Term-Memory(LSTM) networks are newer RNN which have better memory capability. Therefore, LSTM is well developed and suited algorithm to learn from important experiences of LSTM that have very long-short time lags in between

the periods due to their gates usage. The LSTM algorithm is used as building units for the building layers of an RNN, which is then often called an LSTM network. LSTM's forces RNN' store member their important inputs which are further required over a long period of time. This can be performed by LSTM's since they contain their information in a memory which acts as the memory of a computer because the LSTM can CRUD operations on the information from its memory. LSTM will take only the fetch the short-wanted data to predict the future value points.

Basically, there are 3 gates present inside LSTM Model/Idea those are input, forget and output gate. These gates plays a crucial role in determining whether to let new input information in (inputgate), or weather to delete the information because it is not much important(forgetgate) or to let it affect the current time stamp output(output gate). Below we have inserted a RNN typical diagram which consist of 3gates that we discussed above.

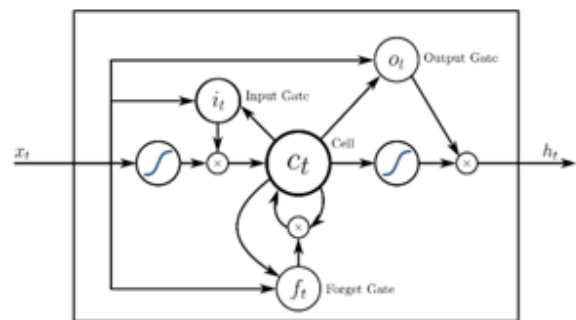


Figure 2: LSTM architecture

ANN Neural Network

ANN is stands for Artificial Neural Network. Artificial neural network are most unique types of ML algorithms that are typical following human brain. The above mentioned is, just same as how the neurons in our brain are fit to mop up from the previous data. identically, the ANN is scope up to study from this and supply responses in the way of forecast, sorting. Artificial neural networks are ahead stastically replica whatever display a difficult of the i/p's and o/p's to come across latest design interrelation in middle. A heterogeneity of quest like speech recognition, picture recognition, device transferring builds this Artificial neural network. A valuable benefit of ANN is the fact that it grasp from the sample datasets. Most continously usage of ANN is that of the accidental function resemblance. By, the help of these kits, Anyone will have fruitful form of report at the answers it defines dispersal. Artificial neural network is also efficient method of picking trail information preferable to the whole data set to give that the outcome of answer. With Artificial neural network anyone will improve current details(data)test technique owing to promote anticipative ability.

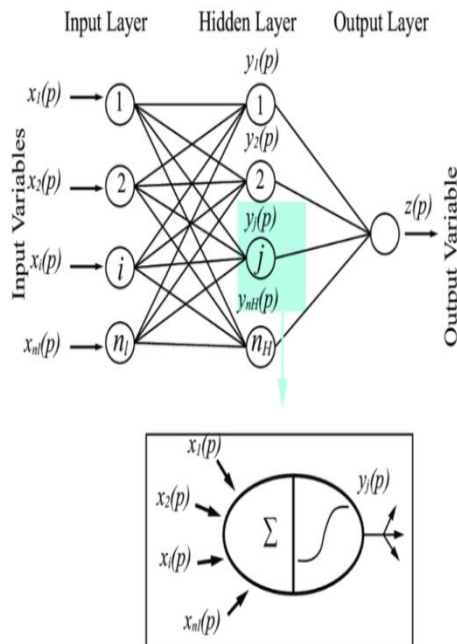


Figure 3: Recurrent Neural Network

Sentimental Analysis

Sentiment Analysis is a new emerging Artificial intelligence innovation which would allow computer to determine mood based on lines of words. It's also known as opinion mining. Sentimental Analysis is used to collect the tweets from the Twitter API which are related to our prediction. To work out this API requires authentication. After using authentication using key, we are able to access through twitter library called Tweepy. The text of each tweets includes URLs, tags and many other symbols, we want filter it out from the original state to get the accurate tweet's sentiment. First it will separate the tweets in to word by words by using python's Natural Language Toolkit(NLP) to remove the words which does not provide any information. After undergoing this process sentiment analysis will predict the mood of the tweet. On the other side we will get the stock value of the product according to the date mentioned in the tweet, it is done with the help of yahoo finance library. The mood of the tweet and the value is combined as the feature set. Using SVM, this feature set will undergo training and predict the value of the stock.

3. Proposed Predicting Module

Our predicting model will consist of two ways of predicting the stock value which allows the user to select any one model at a time according to his need. The first module is the LSTM module and the another one is the sentimental analysis module.

Proposed Architecture

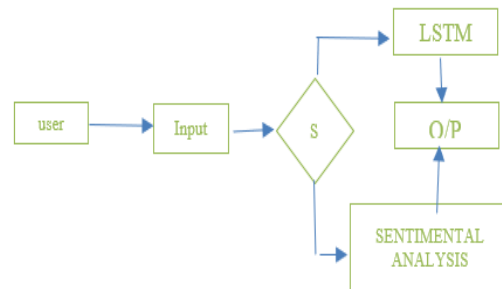


Figure 4: Proposed architecture

In this system the user give throw our interactive web page which is passed to backend as the parameter. There are two inputs given by the user 1. The choice of method he/she needs to use to predict the stock. 2. The choice of stock for which he needs to predict the future then once the back end does the process it again sends the data back to user interface where output is shown to the user

LSTM Module

Formulas used:

$$IT = \sigma(xtUi + HT - Wi)$$

$$FT = \sigma(xtUf + HT - Wf)$$

$$OT = \sigma(xtUo + HT - Wo)$$

$$\sim C = \tanh(xtUg + H - Wg)$$

$$C = \sigma(ft * C - 1 + it * \sim C) \quad HT = \tanh(C) * OT$$

Where IT, FT, and OT are input. It's input j field, FT is neglect field, OT is output field, $\sim C$ is candidate, C is internal memory unit, ht is finding hidden state.

Algorithm

Step1: Start

Step2: Stock data is stored in 3d(N,W,F) Numpy array.

Where:

N-training sequence.

W-sequence length.

F-features of each sequence.

Step3: A network with [I,A,B,L] structure is built, where:

I-input.

A-neuron in next layer.

B-neuron in subsequent layer. L-Single Linear activated layer.

Step4: Train Constructed models on data set.

Step5: Use last step output as prediction of next layer.

Step6: Repeat 4&5 until convergence reached.

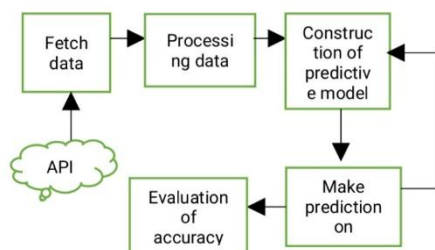
Step 7: Predict test data as input to network.

Step8: Evaluate accuracy.

Step 9: End.

Theory/procedure

If the user selects the LSTM prediction module, he would have to select the stock name for which he would like to predict the stock price and invest on to improve his penny earning to bucks.



The above diagram represents the Architecture of the LSTM model in which each box shows one phase of the process as in of the order. The first box indicates how the historical data of stock price is retrieved from online API. The second step preprocessing include cleaning of data. The columns we have selected after analyzing and filtering the dataset we have generated for this prediction are 'high value' of the stock of the day 'low value' of the stock of the day and the 'closing value' of the stock of the day are assigned to variable regards as the selected data [14][13]. If we plot all these indices on a single graph, The graph that will be obtained for (AAL) stock would be look alike as thefigure-1.This will be our minimized dataset for the prediction. Then we must split this data to train and test the data to train our machine learning bots and to evaluate the bots on the remaining dataastest[2][1].This type of teaching the machine is known as supervised machine learning. [3] In our research we are splitting the dataset in the form of 3:1 ratio the first 75% of the data is to train the data and the remaining 25% is to test the data.Moreover,inthe75% of training the data we will select from the respective filtered fields of the data set to predict the more accurate result so prediction[15]. It is done with the help of LSTM library, since it is a recursive network it will train again the machine bots with 25% of the data that was predicted previously to predict the future value. By following the above-mentioned method for a stock (AAL) [11], we have plotted a graph and stock value which is shown infig-2.The value that we got in sample prediction is much closer to accurate value of the stock market. [12] This show how neural network have achieved a time-series forecasting accurately than other algorithms.

Sentimental Analysis module

Algorithm

Step1: Start

Step2:Provide the authentication for tweeter API.

Step3:Retrieve tweets from Tweeter using this API.

Step4: Once the data is received the it is now Tokenized (i.e.) the Whole sentence is now divided into sperate words. This is done by setting the

delimiter to space. `data.split("\n",text);`

Step5: Cleaning of data: the symbols which does not contribute to sentimental analysis are removed.for symbols in bad_chars : `data=data.replace(symbol,"")`

Step6: Removing of Stop words: not all words are used for predicting so we remove then such as (He,she...). `filtered_sentence = [winstop_words]w for w in word_tokensifnot`

Step7: Classifying each word to calculate its polarity and subjectivity. This is done by using Lexicon which is predefined words like dictionary Now matching the words with Lexicon is done and the polarity is determined.

Step8:End.

Theory

First let us understand why we use sentimental analysis in the project.[12] The stock of a company not only depends on the import and export of a company but also the name and fame of the company in the public view as said earlier. So, we need to analyze what people think about the company or the place of company in public minds. [7] To do so ,need to understand the view of public on company name. This can be done but letting the machine know what people think so we go for a social media platform where people from various places share their view on a topic. In this case we have selected the tweeter because this a social media platform where all the people irrespective of their class share their point.[6]

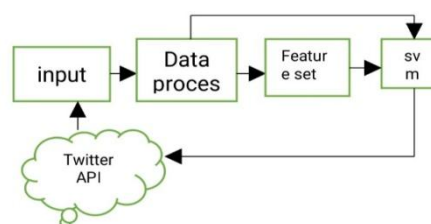


Figure 5: Sentimental Architecture

The process starts by collecting a required amount of live data from the tweeter using the tweeter API. Once the tweets are retrieved successfully they are parsed to understand their subjectivity and polarity [8]. These two parameters are used To understand what may be effect of this information on stock. Later the parsing, made using those parameters where the subjectivity relates to the tweet relation without stock and polarity relates to the reachability of that tweet which can indirectly affect our stock[6]. So,if the polarity is pretty much high than stated level then that tweet is known as the positive tweet and it is likely to increase thestockprices.Ifitisnear0itis known as the neutral tweet which won't be affecting our stock. If its lower than neutral level it is said as negative comment and

may affect in lowering our stock price. This can be used in certain non- predictable events to have a guess of prediction for our safety [10].

4. Result

LSTM model

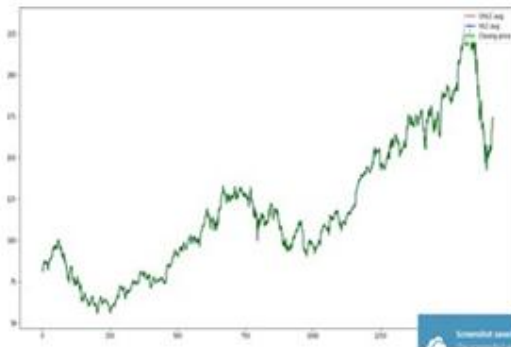
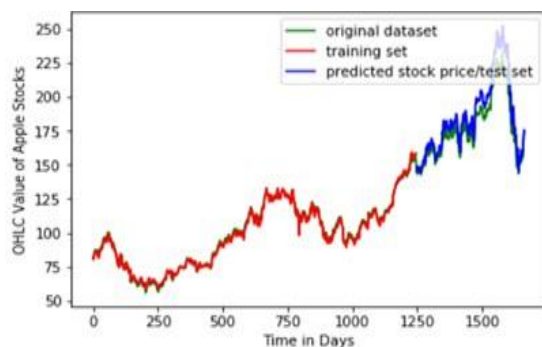


Figure 6: Result after Selection of the Required Parameter



175.58157348632812

Figure 7: Result of final Out put After Prediction

Sentimental Model



Figure 8: Sentimental final output

5. Challenges

The following challenges were addressed during the creation of project:

1. Collecting the appropriate dataset as per the project

requirement.

2. Filtering of the appropriate parameter from the dataset to determine the maximum dependency of the result on parameter.

3. Determining the correct optimization and training algorithm based on the problem statement by understanding the actual need of the work.

4. Finding the required library that would be supportive during the process.

6. Conclusion

Based on this project we conclude by saying that both the methods are efficient and feasible methods for predicting the stock price movement [13]. One model is in the favor of recursive network training and the other model is in the favor of sentiment of tweets [7]. Moreover, our prediction is very inexpensive compared to the ground survey. Our motto in this research is to help the people who are taking risks on investing on stock market.

References

- [1] Sheik Mohammad Idress ,M.Afshar Alam, Parul Agarwal, A Prediction Approach for Stock Market Volatility Based on Time Series Data, 10.1109/ACCESS.2019.2895252, IEEE Access
- [2] Yao Xi Hu1 ,Aoyan Sun2 , (Member, Ieee),Xin Nie3 ,Yuzhu Li4 ,Lian Liu5, An Enhanced LSTM for Trend Following of Time Series, 10.1109/ACCESS.2019.2896621, IEEE Access.
- [3] GUANG LIU1 and XIAOJIE WANG2, A Numerical-based Attention Method for Stock Market Prediction with Dual Information, 10.1109/ACCESS.2018.2886367, IEEE Access.
- [4] Yujie Wang1,2, Hui Liu 1,2, Qiang Guo 1,2, Shenxiang Xie3, And Xiaofeng Zhang4, Stock Volatility Prediction by Hybrid Neural Network,
- [5] An introduction to Indian stock market, <https://www.investopedia.com/articles/stocks/09/indian-stock-market.asp>, date accessed july 2018.
- [6] [Jain, Garima, and Bhawna Mallick. "A study of time series models ARIMA and ETS." (2017).
- [7] Zhang, Q., Li, F., Long, F. and Ling, Q., 2018. Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network. IEEE Access, 6, pp.56984-56994.
- [8] Ashik, A. M., & Kannan, K. S. (2019). Time Series Model for Stock Price Forecasting in India. In Logistics, Supply Chain and Financial Predictive Analytics (pp. 221-231). Springer, Singapore
- [9] L. Chen, Z. Qiao, M. Wang, C. Wanga, R.

- Du, and H. E. Stanley, “Which artificial intelligence algorithm better predicts the chinese stock market?” IEEE Access, 2018.
- [10] X. Li, H. Xie, T.-L. Wong, and F. L. Wang, “Market impact analysis via sentimental transfer learning,” in Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on. IEEE, 2017, pp. 451–452.
- [11] J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models
- [12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Compute.
- [13] Bohn,Tanner. A.“Improving Long term Stock Market Prediction with Text Analysis.” (2017).
- [14] J.R.Thompson, J.R.Wilson and E.P.Fitts, “Analysis of market returns using multiracial time series and agent-based simulation”, in Proceedings of the Winter Simulation Conference (WSC '12). Winter Simulation Conference, Article323, 2012.