Review on Stock Market Forecasting& Analysis-LSTM (Long-Short Term Memory); Holt's Seasonal Method;ANN (Artificial Neural Network); ARIMA (Auto Regressive Integrated Minimum Averag...

Conference Paper · March 2019

CITATIONS
0

READS
158

3 authors:

Anirban Bal
Government College of Engineering and Leather Technology
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Debayan Ganguly
Government College of Engineering and Leather Technology
36 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Kingshuk Chatterjee
Government College of Engineering and Ceramic Technology
40 PUBLICATIONS **35** CITATIONS
SEE PROFILE

Review on Stock Market Forecasting& Analysis

Anirban Bal
Computer Science and Engineering
Government College Of Engineering And Leather Technology, Plot-II, Block – LB, Sector - III, Kolkata – 700106
anirbanrkmission@gmail.com

Debayan Ganguly*
Computer Science and Engineering
Government College Of Engineering And Leather Technology, Plot-II,Block – LB, Sector - III, Kolkata – 700106
debayan3737@gmail.com

Anirban Pal
Computer Science and Engineering
Government College Of Engineering And Leather Technology, Plot-II,Block – LB, Sector - III, Kolkata – 700106
anirbanpal1997@gmail.com

Kingshuk Chatterjee
Computer Science and Engineering
Government College Of Engineering And Ceramic Technology, 73, Abinash Chandra Banerjee Lane,Kolkata-700010
kingshukchaterjee@gmail.com

***Abstract*— In this paper, how a stock can be predicted with time-series forecasting, has been shown. A stock is a time-series data, which has no stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price will maximize investor's gains. This paper proposes a machine learning model to predict stock market price. The proposed algorithm integrates Dense Neural Network (DNN) and Root Mean-Square Error (RMSE). The DNN algorithm is employed to predict the data with hidden feature and RMSE to optimize the model. Proposed model is based on the study of stocks historical data and technical indicators. DNN algorithm selects hidden features to predict the close-price and LSTM (long short term memory) is a model for forecasting time-series data. LSTM predict the next day value depending upon the previous data. MinMaxScaler normalization is done to shorten the range of features between 0 and 1. The proposed model was applied and evaluated using thirteen benchmark financials datasets and compared with artificial neural network with LSTM model. The obtained results showed that the proposed model has better prediction accuracy and the potential of RMSE algorithm in optimizing LSTM. RMSE is used for calculating root mean-square error for determining deviation of predicted value from original one. Another important measurement is sentiment analysis which is done by SOFNN. Since, stock data depends on the sentiment of purchasers and the market sentiment. So, this analysis gives us a more preferred accuracy in stock prediction.**

Keywords*— *LSTM (Long-Short Term Memory); Holt's Seasonal Method;ANN (Artificial Neural Network); ARIMA (Auto Regressive Integrated Minimum Average); PCA; MLP (Multi Layers Perceptron), ;BP-NN (Back Propagation Neural Network); SOFNN (Self Organizing Fuzzy Neural Network), DJIA (Dow Jones Industrial Average)

I. INTRODUCTION

In stock market the forecasting is a necessary way to take decision whether purchasing of the stock is safe or not. There are many public funded markets where the general people invest their money directly. One must be interested to see the future of this stock which is purchased by him or her. Now, stock dataset is a time series dataset which can be inferred by the combination of ANN and time series forecasting [1]. Although the stock market data is not only depend on time but it depends on the previous history of stock data. There may be a repetition of the pattern in the curve for the sequence of the time [2]. We can found this pattern by Holt's linear method or more specifically by 'Holt's Seasonal Method where in a particular season the movement of stock becomes changed and it can be captured by this specific method. Many papers have depicted various kinds of methods for inferring the stock price for future relevance. We need some new model in the field of time series forecasting [3]. Since the data not only depends on the time. So, we can take another field at the independent axis like total transaction, number of shares purchased, or (total number of deliverable quantity/ total no. of purchased quantity). We have also implemented LSTM [4] to predict the next day close price as the structure of previous data impact on the next day forecasting. Many of papers has implemented support vector machine [5], lasso-linear regression [6], LSTM to infer the stock market price for giving some pictorial view. SOFNN is used on the DJIA dataset (Twitter) [16]. Four sentiments have come out such as Calm, Happy, Alert and Kind which steer the movement of stock price. Total 75.6% accuracy is gained by using SOFNN. Another evaluation metric is considered named MAPE.

II. MODELS AND METHODOLOGY

Our proposed models are,

PCA

Combination of features which has greater impact on the label is more useful than a single feature. PCA will take the combination of parameters which have greater impact and will generate the co-variance matrix for determining the number of features.

Scaling the data fields in a lower range is also important [7]

MinMaxScaler

Data normalization is an important preprocessing which can be done using MinMaxScaler normalization. The range will be between 0 and 1. Since, many features have too large values (Turnover), and some of them have too small value (%Deli. / traded Qty.), they must be taken in a small range if we want to use ANN (which is most effective in case of long time forecasting) or RNN.

Selecting correct feature is important too [3]

Feature Selection

We have to select those features, which uniquely impact on stock and this is a unique or combination of unique features in the dataset. Since, high-low spreading, No. of traded quantity, %Deliverable quantity/ Traded quantity, No. of shares, Turnover have impact on the closing price directly. Investors' sentiment is another feature to predict some uncertainty in the stock prediction field.

Un-bias data selection is one of the prior processes to select the dataset.

Un-biasness

Un-biasness in the sample is not good for making a prediction. If different type of trends is followed in the section of test data, the forecasting may be wrong in case of ARIMA or time series forecasting method where the prediction is made from the previous trends. New trend is always un-recognizable in case of trends reading method. So, random sampled data will be taken for getting a better result in any method as well as ARIMA.

Selection of correct model is also important [7]

Model Selection

ARIMA [8] will be selected as a method for forecasting short range data rather than long range because short range testing data may have low number of variation than long data. So, repetition of trend is merely impossible in case of long time forecasting.

Many papers have implemented various regression models, learning, pre-processing for achieving the best result in case of finding the forecasted data. Those models are linear model, lasso-linear model, and support vector machine, LSTM, ARIMA, and Stochastic Gradient Descent etc. [9] linear model was less effective in case of polynomial curve detection. So, it is less sensitive to the normalization method. For this reason the window size is varied and the re-evaluation is done to predict more effectively.

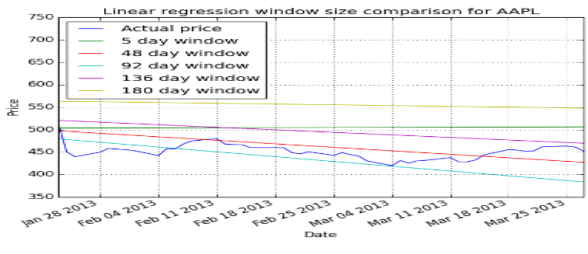

Figure 1: Linear regression window size comparison for AAPL [6]

As per source document the SGD can be used on that dataset which is larger than 10000. So, the share price of that company can't be predicted using SGD (Source: Stock Market Price Prediction Using Linear and Polynomial Regression Models)

In case of SVR the result is better than others. But the selection of kernel made differences.

Those differences are,

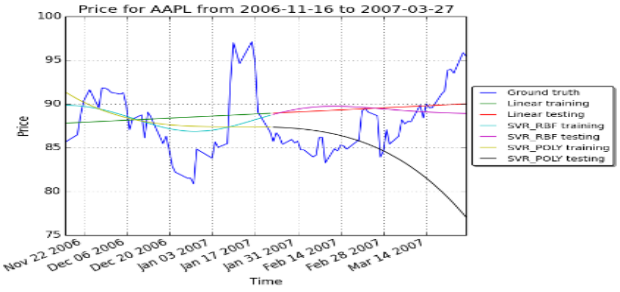

Figure 2: Price for AAPL from 2006-11-16 to 2007-03-27[6]

Since, error is a determining factor in share price forecasting. The following image will help a lot,

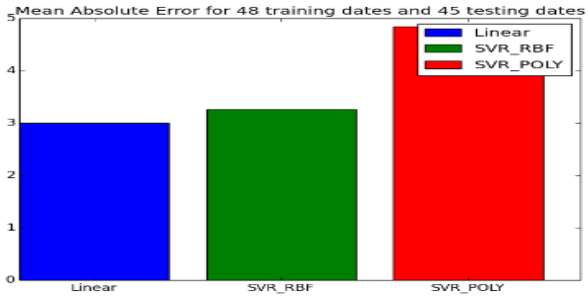

Figure 3: Mean absolute error for 48 training dates and 45 testing dates [6]

For the impact of many features stock prices can be forecasted by using LSTM.

It can tackle many features, it is more accurate and reliable since it considers many features and decides according the relation and impact of them on the label. RNN has some disadvantages too. Market condition and environmental influences cannot be considered. We have to consider each feature of stock as a time series variant rather than direct variants. ANN has a great disadvantages which are, finding the architecture of the model in ANN is harder and we have to go forward blindly (on the basis of loss calculation) in case of tuning the ANN model. So, finding any statistical structure is

somehow easy with the help of visualization method rather than implementing ANN. [10]

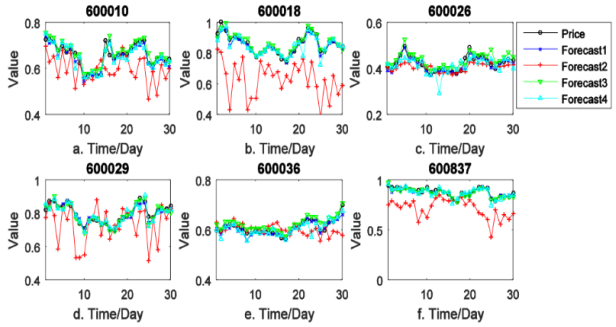

Figure 4: The original series and the forecasting series with different methods [11]

The RME of corresponding result is,

Ticker	MSE_P	MSE_S	MSE_M	MSE_R	B_{PM}	B_{PR}
600010	0.000441	0.00690	0.000602	0.000561	26.80%	21.47%
600018	0.000357	0.06542	0.000371	0.001044	3.773%	65.76%
600026	0.000186	0.00090	0.000521	0.000926	64.37%	79.95%
600029	0.000260	0.0143	0.000379	0.000648	31.37%	59.83%
600036	0.000141	0.00157	0.000275	0.000321	48.68%	55.98%
600837	0.000332	0.0366	0.000344	0.000609	3.528%	45.47%

Table 1: The MSE of different models and B value [12]

Using artificial neural network models in stock market index prediction [13]

Here multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks which use generalized autoregressive conditional heteroscedasticity to extract new input variables are used. The error term is most important in term of finding best model. So, RMSE is the best for this purpose.

Uniqueness of this study is that here comparison among different ANN models are used.

DAN2 architecture worked more like a statistical method rather than an ANN. From overall results, we can classical ANN model MLP works the best in forecasting time series. Hybrid methods failed to improve the forecast results.

Predicting stock market index using fusion of machine learning 4 techniques [14]

Here a fusion approach is used. In first stage SVR is used. Then in 2nd stage ANN, RF, SVR are used interchangeably.

The 1st stage of this approach processes the data for the prediction models in the second stage.

In this study SVR–ANN model works the best. For long time forecasting the ANN model works well and for short time forecasting SVM is better.

Empirical analysis

Stock market prediction via extreme learning machine [15]:

Here they used ELM to build a model that can take news & market prices as input feature and predict stock short-term price movements with good accuracy and fast speed.

They compared the performance of ELM with SVM and BP-NN on one year H-share news articles and intraday tick prices.

Stock price prediction using the ARIMA model

From the results, we observed ARIMA performs better for short-term prediction compared with other existing models like ANN, SVR etc.

In ARIMA the short time forecasting results well because it forecast from the previous data trend. For this reason ARIMA is used for short time data.

Sentiment analysis using SOFNN [16]

The Self Organizing Fuzzy Neural Network (SOFNN) is a five layer fuzzy neural network which uses ellipsoidal basis function (EBF) neurons consisting of a center vector and a width vector. This is used to find the non linear trends which are typically usual for the stock market data.

Steps under this process are,

Stock data hugely depends on the behavioral economics like purchasers' and markets' sentiment.

The Efficient Market Hypothesis (EMH)[16] states that stock market prices are largely driven by new information and follow a random walk pattern. Though this hypothesis is widely accepted by the research community as a central paradigm governing the markets in general, several people have attempted to extract patterns in the way stock markets behave and respond to external stimuli.

Preprocessing is done like null filling with average value of its neighbors, local obligation overcoming by step up/down and removing highly varied data.

Wordlist is done by POMS. 65 questions on current sentiment are asked for normalizing the sentiment data. Sentiwordnet and a standard Thesaurus is used to asking the questions.

Twitter filtering is done based on those tweets which are more likely to express a feeling, i.e. we consider only those tweets which contain the words "feel", "makes me", "I'm" or "I am" in them to save computation time to make the solution real time feasible.

Word tagging and opinion making is done by the daily score comparison. The required formula is,

Score of a word = #of time the word matches tweets in a day / #of total matches of all words.

Score mapping is done on the basis of the answer on the six states of POMS

Table 2: p-values obtained using Granger causality analysis with different lags (in days)[16]

Lag	*Calm*	*Happy*	*Alert*	*Kind*
1	*0.0207*	*0.4501*	*0.0345*	*0.0775*
2	*0.0336*	*0.1849*	*0.1063*	*0.1038*
3	*0.0106*	*0.0658*	*0.1679*	*0.1123*
4	*0.0069*	*0.0682*	*0.3257*	*0.1810*
5	*0.0100*	*0.0798*	*0.1151*	*0.1157*

Algorithm	Evaluation	ID	I_{CD}	I_{CHD}	I_{CAD}	I_{CKD}	I_{CHAD}	I_{CHKD}
Linear Regression	MAPE	7.28%	7.26%	7.66%	7.05%	7.43%	7.57%	7.78%
	Direction	64.44%	64.44%	71.11%	64.44%	64.44%	68.89%	71.11%
Logistic Regression	Direction	60%	60%	60%	60%	60%	60%	60%
SVM	Direction	59.75%	59.75%	59.75%	59.75%	59.75%	59.75%	59.75%
SOFNN	MAPE	9.71%	9.66%	11.03%	9.22%	11%	10.52%	11.78%
	Direction	64.44%	71.11%	75.56%	68.89%	73.33%	73.33%	73.33%

In this paper, the closing price parameter is chosen for time series forecasting.

So, it shows all the activities of a trading day.

III. CONCLUSION

This paper is presenting the recent work on the stock market analysis and forecasting which is presented for decoding the volatility and versatility of equity share movement. Many models are implemented on the basis of various learning algorithms. Many forecasting method is related to the time series prediction which is made here. We have summarized the total method which is giving a prediction of future data from the current day on which we are present.

REFERENCES

[1] C. Narendra Babu, Eswara Reddy, A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data, doi: https://doi.org/10.1016/j.asoc.2014.05.028 , (5 May, 2014).

[2] JoyjitDhar, Investment Management of Mutual Funds: Evidence of Timing and Selectivity from India during 1997-2003

[3] Huang, C.-L., & Tsai, C.-Y. A hybrid SOFM–SVR with a filter-based feature selection for stock market forecasting is done. Expert Systems with Applications, 36, 1529–1539 A.N. (2009).

[4] Menezes, L. M., &Nikolaev, N. Y., Forecasting with genetically programmed polynomial neural networks is done. International Journal of Forecasting, 22, 249–265, (2006).

[5] PhayungMeesad, Risul Islam Rasel,Predicting Stock Market Price Using Support Vector Regression , DOI: 10.1109/ICIEV.2013.6572570. (2013).

[6] Lucas Nunno ,Stock Market Price Prediction Using Linear and Polynomial Regression Models, University of New Mexico, Computer Science Department, Albuquerque, New Mexico, United States, lnunno@cs.unm.edu. (2014).

[7] QunZhuge, Lingyu Xu and Gaowei Zhang ,LSTM Neural Network with Emotional Analysis for Prediction of Stock Price, (2017).

[8] A. Zapranis, G. Francis, Stock performance modeling using neural networks: a comparative study with regression models, Neural Networks, 375–388 (1994).

[9] Gooijer, J. G. D., & Hyndman, R. J. 25 years of time series forecasting. International Journal of Forecasting, 22, 443–473.(2006).

[10] Leu, Y., Lee, C., &Jou, Y. ,A distance-based fuzzy time series model for exchange rates forecasting. Expert Systems with Applications, 36, 8107–8114, (2009).

[11] Atsalakis, G. S., &Valavanis, K. P., Forecasting stock market short-term trends using a neuro-fuzzy based methodology is done. Expert Systems with Applications, 36, 10696–10707, (2009).

[12] K.H Huarng. T.H.K. Yu, 'The application of neural networks to forecast fuzzy time series', Phys. A 336 481-491 (2006)

[13] ErkamGuresen, GulgunKayakutlu ,Tugrul U. Daim, Using artificial neural network models in stock market index prediction, doi:10.1016/j.eswa.2011.02.068 , (2011).

[14] Xiaodong Li, HaoranXie,RanWang,YiCai,Jingjing Cao,Feng Wang,Huaqing Min,Xiaotie Deng,Empirical analysis: stock market prediction via extreme learning machine,Neural Compute & Application DOI 10.1007/s00521-014-1550-z, (2014).

[15] Jigar Patel, Sahil Shah, Priyank Thakkar , K. Kotecha, Predicting stock market index using fusion of machine learning 4 techniques,https://doi.org/10.1016/j.eswa.2014.10.031 , (2014).

[16] Anshul Mittal, Arpit Goel, Stock Prediction Using Twitter Sentiment Analysis. (2013)