Prophecy of Stock Price of Bharat Immunological & Biological Corporation Ltd using Hybrid ANN and PSO Model

Mohammed Siddique¹ and Siba Prasad Mishra²*

¹Department of Mathematics, Centurion University of Technology and Management, Odisha, India.
²Department of Civil Engg., Centurion University of Technology and Management, Odisha, India.

Authors’ contributions

This work was carried out in collaboration between both authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JEMT/2021/v27i430338

Editors:
(1) Dr. Olusegun Felix Ayadi, Texas Southern University, USA.
(2) John M. Polimeni, Albany College of Pharmacy & Health Sciences, United States.

Reviewers:
(1) Tarjei Kristiansen, Denmark.
(2) Damian Sambuo, Mushi Co-operative University (McCU), Tanzania.
(3) Zohanto Widyantoko, Indonesia.
(4) Daniel Febrian Sengkey, Universitas Sam Ratulangi, Indonesia.
(5) Manoj Kumar Kowar, India.

Complete Peer review History: http://www.sdiarticle4.com/review-history/70457

Received 01 May 2021
Accepted 06 July 2021
Published 09 July 2021

ABSTRACT

The prediction of the time series has always attracted much interest from investors and researchers to evaluate financial risk. Stock market movements are extremely complex and are influenced by different factors. Hence it is very important to find the most important factors for the stock market. But the high level of noise and complexity of the financial data makes this job very difficult. Many authors have already used artificial neural network for this kind of forecasting tasks, but hybridization model of artificial neural network is considered to be widely used and better performing forecasting model among others. The dormant high noises data mess up the performance, so to enhance the prediction accuracy. We considered a set of seven technical attribute of stock market to perform the hybrid model of Artificial Neural Network (ANN) and Particle Swarm Optimization algorithms. The efficiency of the proposed method is measured by the stock price of Bharat Immunological & Biological Corporation Ltd with 3945 number of daily transactional data. Empirical prediction analysis shows that the proposed model enhances the performance in comparison to simple ANN model.
1. INTRODUCTION

Exploration of the stock markets has always been an indispensable fragment financial sector of any country. Many investors currently rely on smart trading systems to forecast stock market price, based on different conditions. Precision of these forecast systems with minimum risk factors is important for better investment decisions. The stock price prediction has been useful to individuals as well as institutional investors. The stock data is most difficult as it is complex, non-linear and not every parameter. Investors therefore find it very difficult to predict financial stock market without studying their patterns. The selection and execution of an appropriate forecasting method plays a vital role to most of the financial traders. The organizational and monetary permanence of an organization depends on the accurateness of the prediction as such information determines to make key decisions in the areas of possessions, acquisition, advertising, planning and development of any organization and firm. The objective of this work is aligned towards the development of capable machine learning based hybrid model which forecasted the higher accuracy of stock market Armano et al. [1], Budiharto [2].

Technological analysis is an admired approach to study the stock market analysis. To forecast future trends or prices, Researchers use a range of machine learning and intelligent artificial approaches. Many artificial neural network (ANN), support vector machine (SVM), and logistic regression (LR) employed for such type of prediction responsibilities. The ANN approach to attack the logistic is considered to be one among the best performing modus operandi where the regularization parameter is properly initialized. We used the artificial neural network and particle swarm optimization technique for forecasting the stock price of Bharat Immunological & Biological Corporation Ltd. (BIBCOL) Technical metrics are derived from the historical trading data used in this study.

The BIBCOL is a biotechnological Government of India concern in Uttar Pradesh based in Uttar Pradesh India. The company is manufacturing pharmaceuticals products like Oral Polio Vaccines, Zinc Tablet & Diarrheal Management Kit etc. which has high demand as ariptic drug during the present apocalyptic pandemic situation in India. So it was considered to make stock price prediction studies.

2. LITERATURE REVIEW

The literature review on the topic ANN has extended significantly. Initially the authors Reid [3], Bates et al. [4], Clemen [5], Rajesh et al. [6] and Farahani et al. [7] were broadly discussed about ANN and interpreted the source of information in this field. Wedding [8], Kumar Chandra [9], proposed a hybrid model of ANN with the radial basis function. Luxhoj, et al. [10] proposed an econometric model of ANN to forecast the sales price. Pelikan [11], Ginzburg [12], and Kumar et al. [13] were developed a feed forward neural networks to forecast the power consumption and enhanced the prediction precision of time series data. Tsaih [14] developed a novel ANN model by integrating the concept of artificial intelligence (AI) to predict the index of American stock market. Medeiros [15] constructed a hybrid model of ANN with smooth transition auto regressive model to forecast time series data. In current years, many hybrid prediction models have been proposed by incorporating ARIMA with ANN to archive the better performance. Pai et al. [15] incorporated SVR along with ARIMA for forecasting stock price. Armano [16] proposed a novel hybrid model by incorporating GA with ANN and applied to predict stock index. Lean Y, et al. [17] had presented a nonlinear forecasting model by assembling ANN with Generalized Linear Auto Regression (GLAR) to predict accuracy in stock market. Chen et al. [18] proposed a hybrid forecasting model by the combination SARIMA with SVR. The researchers, Khashei, et al. [19] have developed a hybrid model using the core principles of artificial neural networks to address nerve limitations in case of incomplete data sets. Khashei et al. [20] Siddique et al. [21] proposed hybrid model of ANN-ARIMA and got more precise prediction and outputs are better than the results obtained from artificial neural networks. Superiority of hybrid model ANN with PSO has been claiming by various authors over other models Armaghan et al. [22], Pranav et al. [23], Gourav Ku. et al. [24].

Why Hybrid ANN (PSO) model: The notable task for investors is to handpick farms capitalize by analysing financial data of stock price which is a tedious assignment. Financial issues of
exchange rate, and share value in stock market is an exciting innovative research arena. The stock price prediction is presented in the current research during the Pandemic Covid-19 is very challenging particularly in the field of Vaccine and allied health securities. Many researcher have tried with many models like artificial neural networks (ANNs), support vector machines [3], genetic algorithms (GA) and has opined in favour of Artificial Neural Network (ANN) and modified hybrid Particle Swarm Optimization (PSO) [25] Burges, [26], Majhi et al. [27], Gupta et al. [28] Ahamad et al. [29], Siddique et al. [30], Farahani et al. [7].

3. METHODOLOGY USED

3.1 Artificial Neural Network (ANN)

ANN is a parallel processing computational model, motivated by the efficient work process of genetic neurons in which mammalian brain progress information represent to it. ANN is a compilation of mathematical models and imitates some of the empirical phenomenon in a genetic nervous system, most prominently adaptive natural learning. ANN is a computational method which is encouraged by working principle of human brain and is constructed to approximate functions of large number of inputs and is generally unknowns. It is ordinarily represented as a group of interconnected "neurons" which can compute values from inputs and that is why these are termed as neural networks. An ANN consists; sets of adaptive weights that are tuned by learning algorithm (it is similar to β’s in multiple linear regressions) and some functions that can best adapt to these weights and give the desired estimates, Zhang, et al. [31]. The importance of NN in mathematical modelling is its non-linear nature which is very much befitting to establish complicated relationships between inputs and outputs. Each neuron in the network mapping is associated with some weight. Every input neuron is multiplied by the corresponding assigned weights. Bias is added to improve the network performance. It adjusted the output neurons with the weighted sum input. After that sum of above products are passes through an activation function to compute the result. Sigmoid function is used for activation function. Our aim is to compute the output neurons for every set of input neurons.

3.2 Particle Swarm Optimization (PSO)

PSO is one among the leading meta-heuristic optimization methods which is motivated by birds and fishes co-ordinated, collective social behaviour. The basic concept of PSO was developed by Kennedy et al. [32], ARIashidi et al. [33], Wang et al. [34]. The divergent social behaviour of flocking of avifauna and fish schooling has inspired them to develop the algorithm. When searching for food, the basic concept that the model was based on is that the birds are either dispersed randomly or disperse together before vacating the area for venturing another place where they could find satisfactory, large in quantity and good quality of sustenance source that can explore. Amidst the search process, there are few birds that always transmit in advance during transit and provide information about the place where ample, and quality food is available to the other flocking birds. They land the forecasted place of adequately available food in quantity and quality because of this foresight and the knowledge update which is shared between them. This behaviour is delimited by a populous or individual particle called a swarm. The swarm concept can be developed used by numerous iteration finally resulted in moving to the optimal solution in their area. The particles are connected to the current position, and the velocity, by which they can have the finest knowledge of the top local location (blest) that each swarm has found and the finest global position (gbest) achieved by all particles. For each iterative stage the particles travel by applying their corresponding velocity from the current position. In the previous iteration, the vector parameters like the direction and the magnitude of the velocity is resolved by simulating momentum, the velocity, and the relative position to the (lobest) and (gbest). Motion mechanism which was dynamically adapted the velocity and location of the swarm particles evolved in each generation can be formulated as

\[ v_j^{p+1} = w \times v_j^p + c_1 \times rand(0,1) \times (lbest_j^p - x_j^p) + c_2 \times rand(0,1) \times (gbest^p - x_j^p) \]

\[ x_j^{p+1} = x_j^p + v_j^{p+1} \]

Where; \( v_j^{p+1} \) and \( v_j^p \) stands for the velocity of \( j^{th} \) particle in \((p+1)^{th}\), and \( p^{th}\) iteration, respectively. “w” represents the coefficient of initial weight; “c_1” represents the personal factor and \( c_2 \) represents the social learning factor. “rand” is a random number in \((0, 1)\). \( x_j^{p+1} \) & \( x_j^p \) denotes the location of \( j^{th} \) particle found...
4.2 Pre-processing

In the present study, the data used were of very large values. So initially we have to normalize the big data. The data standardization was introduced to condense the range and to counteract the superiority of data ranging is done by prioritizing the greater numerical values over smaller numerical ranges. The data standardisation reduces the data's gross effects. It maximizes its integrity as that helps to achieve improved accuracy by various algorithms. The transfer of the range of original data was done to another scale. The data was moved to another scale in this phase of normalization. By scale transfer, it was possible to moderate its range and continued to transfer the data more rapidly. Present hybrid model was the set of data which was structured using the equation.

\[ NV_k = \frac{A_k - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} \text{, for } k = 1, 2, 3, ..., l \]

Where “Ak” signifies the original price of the “k”th structures, “l” denotes the total number of the processed data,” A_{max}” is the uppermost value, and “A_{min}” is the lowermost value in the entire time series data set. “NV_k” is the analogous normalized value as referred to. The result of normalization of the data series was lying within the closed range [0.0, 1.0].

4.3 Design and Implementation

Artificial neural networks (ANN) can estimate a huge data set with high prediction accuracy since the input information is processed parallel. The network model does not require any prior assumption where as it is mainly build up by the characteristics of the data. We design the ANN architecture comprises of three layers, such as input layer, followed by the hidden layer, and finally the output layer. The output layer processed from the hidden layer. The activation function helps to convert input signal into output signal. Here we used the sigmoid activation function to process the hidden layer. This model was characterized by the above three network layers connected through finite directed graph.

The relationship among the output \((y_n)\) and the input nodes \((y_{n-1}, y_{n-2}, ..., y_{n-p})\) is representing mathematically as follows

\[ y_n = w_0 + \sum_{j=1}^{p} w_j g\left(w_0 + \sum_{i=1}^{l} w_{ij} y_{n-i}\right) (1) \]

Where \(w_0\) is the bias node used to adjust the output along with the weighted sum of the inputs to the neuron and \(g(\cdot)\) is the activation function, which processed the received information. \(w_{ij}(i = 0, 1, 2, 3, ..., P, j = 1, 2, 3, ..., Q)\), and \(w_j (j = 1, 2, 3, ..., Q)\) are the weight allocated to the input nodes. The ‘P’ and ‘Q’ signify the integral value of the input nodes, and hidden nodes respectively of the designed network. The sigmoidal function \(Sig(x) = \frac{1}{1 + e^{-x}}\) is used to determine the output of the neural network. Sigmoid function is smooth and bounded function which gives real valued output between 0 and 1. The ANN model is designed by the equation (1) mapped from the precedent observation to the upcoming target value \(y_k = f(y_{k-1}, y_{k-2}, y_{k-3} ..., y_{k-p}, W)\), where \(W\) is the vector representation of all the weights and \(f(y_{k-1}, y_{k-2}, y_{k-3} ..., y_{k-p}, W)\) is the function of connective weights of network structure. Hidden layer Q depends upon the input data and there is no specified rule is to assign the parameters weight.
Neural network architecture is governed by input, output and one hidden layer. The process of setting the parameters for the neural network to mimic a particular behaviour is called the learning algorithm. It can be defined as a set of rules for finding optimum weight and bias values which optimize the performance of the neural network. There are different techniques that are used to find suitable values of ANN weights and biases depending on the form of learning the weights of the neural network are adjusted to maximize the output according to specific criteria. The Sigmoid activation function is used to approximate continuous functions as it takes any real valued input and returns a bounded output from 0 to 1. The function is used as a transfer function in this analysis. To build an ANN, we need to determine how many inputs, how many hidden layers and how many hidden neurons there are. Weights and bias are initially set at random and are modified by a learning algorithm during the training phase. After defining the number of input neurons, we need to optimize the weight and bias of both hidden and output layer of ANN using the technique of optimization algorithm, the Particle Swarm Optimization (PSO). PSO is one metaheuristic information sharing the innovative technique where the particles bring up-to-date themselves with the internal velocity. PSO requires few parameters to adjust and gives the optimal output with less number of iterations .PSO update the weights and bias until to get the optimum criteria.

The proposed hybrid model (ANN-PSO) has been implemented in Mat Lab r2013a with the help of library file FANN (Fast Artificial Neural Network) This hybrid model was designed with ANN at the core along with PSO to optimize its parameters. The system that was used for Intel® Corei3-4005U 1.7GHz 4 GB RAM implementation.

5. METRIC AND ITS DEFINITION

5.1 Estimation of Errors

We have used three common statistical metrics for output assessment of the proposed hybrid model. The main objective of applying the modern developed algorithm was to minimize the predictive errors in order to obtain high precision in the proposed model by the values of performance matrices, (Table 2).

Table 1. Features selected of Bharat Immunological & Biological Corporation Ltd. under the study of ANN and PSO

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Open Price</td>
<td>Highest Price</td>
<td>Lowest Price</td>
<td>Closing Price</td>
<td>Number of Shares</td>
<td>Number of Trades</td>
<td>Turnover</td>
</tr>
</tbody>
</table>

Fig. 1. Flow chart of ANN-PSO Mechanism
6. RESULTS ANALYSIS

The performance of present projected hybrid model (ANN-PSO) is matched with standard ANN model is designed with the core concept of ANN with single hidden layer and PSO optimized the assigned weight and bias to minimize the prediction error. The data set is divided in training and testing phase as mention below:

Original dataset size: (3945, 7)
Training dataset size: (3156, 5) (80% of original dataset)
Testing dataset size: (789, 5) (20% of original dataset)

Table 2. The performance measuring matrices like MSE, MAE, RMSE for the hybrid model

<table>
<thead>
<tr>
<th>SL</th>
<th>Metric</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MSE (mean square error)</td>
<td>[ \frac{1}{T} \sum_{i=1}^{T} (y_i - d_i)^2 ]</td>
</tr>
<tr>
<td>2</td>
<td>MAE (mean absolute error)</td>
<td>[ \frac{1}{T} \sum_{i=1}^{T}</td>
</tr>
<tr>
<td>3</td>
<td>RMSE (Root mean square error)</td>
<td>[ \sqrt{\frac{1}{T} \sum_{i=1}^{T} (y_i - d_i)^2} ]</td>
</tr>
</tbody>
</table>

Where, \( T \) is the complete evaluation information, \( d_i \) is the initial value and \( y_i \) is the integral value acquired by the forecasting process.

Table 2. Performance of ANN-PSO Models on the data set of Bharat Immunological & Biological Corporation Ltd

<table>
<thead>
<tr>
<th></th>
<th>ANN Model</th>
<th>ANN-PSO Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>MSE 3.5938888210</td>
<td>0.2557190433158905</td>
</tr>
<tr>
<td></td>
<td>MAE 0.856451993</td>
<td>0.36293157199448567</td>
</tr>
<tr>
<td></td>
<td>RMSE 1.895755314</td>
<td>0.5056867047054832</td>
</tr>
<tr>
<td>Testing</td>
<td>MSE 4.173893917</td>
<td>0.27902256186002816</td>
</tr>
<tr>
<td></td>
<td>MAE 1.202996162</td>
<td>0.3749166291345631</td>
</tr>
<tr>
<td></td>
<td>RMSE 2.043010993</td>
<td>0.528225625436927</td>
</tr>
</tbody>
</table>

In this analysis, the errors measured in the training phase with MSE, MAE and RMSE in the Bharat Immunological & Biological Corporation Ltd data sets were 0.2557190433158905, 0.36293157199448567, and 0.5056867047054832 respectively. The errors obtained in the test phase were 0.27902256186002816, 0.3749166291345631, & 0.528225625436927 respectively.

Fig. 2. Actual Vs Predicted value of Bharat Immunological & Biological Corporation Ltd in training Phase
7. CONCLUSION

In this chapter, an efficient hybrid model using ANN and PSO for forecasting stock price has been proposed. The hybrid model, ANN-PSO comprises of two innovative techniques, which can be applied in the predicting issues of the following day open stock price. The results suggested that the model was simultaneously suitable for the analysis and also the output can be considered from point of view of appropriate implementation. The dataset comprised of 35 attributes which included seven critical features (Table-1) over the past consecutive five days lagging data for open price forecast. The test results from the empirical study showed root mean square error (RMSE) of 0.50 and 0.52 (approx.) in the training and testing phase respectively. The experiment result is tested and verified in Mat Lab r2013a in normal system configuration. This proposed model can be used for taking better decision and more accurate predictions for financial investors on daily stock market forecasting. From the application point of view, we are quite hopeful that our proposed model (ANN-PSO) will be of great help to forecast not only the stock price but also every aspect in the financial domain.

DISCLAIMER

The company Bharat Immunological & Biological Corporation Ltd used for this research is commonly and predominantly selected in our area of research and country. There is absolutely no conflict of interest between the authors and company because we do not intend to use this company as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the company rather it was funded by personal efforts of the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCE

5. Clemen, R. Combining forecasts: A review and annotated bibliography with


© 2021 Siddique and Mishra; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/70457