Stock Market Prediction Framework Based on the Fruit Fly Optimization Method

Parag Rastogi

*Abstract***:** *Stock market prediction helps investors in decision-making process of investment to achieve profit. Recently, the deep learning method shows the significant performance in the stock market prediction. These deep learning models have the drawback of overfitting problems when it processing number of features. In this research, the fruit fly optimization method has been proposed for the feature reduction process in the stock market prediction. The fruit fly method has the advantages of simple computation processes and less number of parameter for tuning. The fruit fly method selects more relevant features to reduce the overfitting problem in the Long Short Term Memory (LSTM) classifiers. The Nifty 50 and S&P 500 data were applied to test the efficiency of the proposed model. The obtained result shows that the fruit fly method based framework achieved more efficiency than other techniques. The fruit fly based framework has 0.426 of Mean Square Error (MSE) and the existing firefly method has 0.621 MSE.*

Keywords **:** *Fruit Fly Method, Long Short Term Memory (LSTM), Mean Square Error (MSE), Overfitting, and Stock market prediction.*

I. INTRODUCTION

Stock market forecasting is an interesting topic in financial time series analysis and has attracted many researchers. Prediction of stock market prices has been a major challenge for investors due to many uncertain parameters [1]. Neural Network has been used for non-linear analysis in many areas such as financial securities, Signal Processing and Pattern Recognition. Moreover, this method has been widely used in many applications like stock market forecasting, document classification and so on due to its advantages [2]. Furthermore, many optimization methods such as Genetic Algorithm (GA), Firefly Algorithm (FA), and others have been used in feature selection with Artificial Neural Network (ANN) to increases efficiency of stock market forecasting [3]. Stock market forecasting is complex and non-linear in nature, with complexity is created by the correlation of market behavior and investment psychology. Many machine learning techniques such as ANN and Support Vector Machine (SVM) have been used to address this problem and improve the performance [4].

Deep Learning method has the huge success in the image classification model and has attracted considerable attention. Deep learning with the advancement of learning capability is expanding to various range of applications [5]. Technical analysis is the process of extracting external technical indicators in the stock and trade market based on the numerical indicators. Fundamental analysis method is an analysis of the financial status of the company and both of this analysis has been applied in the stock market forecasting [6]. Stock analysts believes that price is able to explain the stock market movements and they usually consider the mathematical indicators for prediction. Many existing methods have been applied in stock market prediction and these methods have the drawbacks of overfitting and lower performance [7, 8]. This research applies the fruit fly method for the feature reduction in the stock market prediction. The analysis shows that fruit fly based architecture has more efficiency than other feature reduction method compared.

The paper is organized as follows, the stock market prediction recent researches were surveyed in section 2, the explanation about the fruit fly based method is provided in section 3, the design and parameter settings are detailed in section 4, the experimental result of fruit fly based architecture are explained in section 5 and the conclusion is given in section 6.

II. LITERATURE SURVEY

Stock market prediction is essential for the investors to reduce the risk of decision making in the stock investment. The researches involved in applying stock market forecasting were surveyed in this section.

Baek, et al. [9] developed a data augmentation method for improving the prediction performance of stock market prediction. The developed augmentation method reduces the overfitting problem in the LSTM model. The developed method efficiency is measured in the Chinese and American stock data. The data augmentation and LSTM method have more efficiency in the stock market prediction. This method uses the closing stock index and relevant stock for the prediction. However, the use of more technical indices helps to improve stock market efficiency.

Chen and Hao, [10] proposed a hybrid model of feature weighted SVM and feature weighted K-Nearest Neighbor (KNN) for effective prediction of stock value. The featured weighted SVM is applied for the data classification that assigns the weight values for different features based on the classification importance and each parameter is measured based on the information gain.

Retrieval Number: D75900494202020©BEIESP
DOI: 10.35940/ijeat.D7590.049420
Journal Website: www.ijeat.org

957

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication
© Copyright: All rights reserved.

Stock Market Prediction Framework Based on the Fruit Fly Optimization Method

The KNN is used to forecast the future stock value by processing weighted KNN. The Chinese stock market data has been used to estimate the effectiveness of the developed model. Efficient correlation weight method can be used to increase the efficiency of the model.

Das, et al. [11] applied a feature optimization method based on the firefly algorithm for forecasting of stock market indices. The technical indices and statistical measures were used to improve the model efficiency. The feature reduction process is carried out for the dataset before applying in the prediction model. Some prediction models were applied to analyze the developed method where both statistical and optimized feature reduction methods were used. The statistical methods such as Factor Analysis and Principal Component Analysis (PCA) and optimization method such as Firefly Optimization and Genetic Algorithm were used. The experimental analysis shows that the firefly method outperforms other methods in the feature reduction method. The LSTM model can be used to improve the efficiency of the developed method.

Ramezanian, et al. [12] developed a framework of Genetic Network Programming (GNP) with reinforcement learning and Multi-Layer Perceptron (MLP) for stock market prediction. The model was developed to forecast one-day return of the stock value. The 9 stock data with 5 technical indicators in 3 window is applied to estimate the model efficiency. The experimental result show that proposed model has more efficiency in stock market prediction and the model has an overfitting problem as relevant feature is required to be selected.

Long, et al. [13] designed a feature selection method based on the end-to-end model from the Multi-Feature Neural Network (MFNN) for stock market prediction. The recurrent and convolutional models are integrated for multi-filters structure to provide the essential features for market values. The Chinese stock market data were applied to test the efficiency of MFNN for stock market prediction. The analysis shows that the developed model has more performance. The developed method has the overfitting problem and computational time is high.

III. PROPOSED METHOD

Stock market forecasting is one of the important tool for investors to help in decision- making process. Many studies have been conducted to increase the performance of the stock market prediction. In this research, a framework has been proposed based on the fruit fly method and LSTM. The technical indicators were extracted from the data to increase the efficiency of the method. The fruit fly method has been used for the feature reduction technique. The relevant features are provided for the LSTM architecture for the stock market prediction. The block diagram of the proposed fruit fly with LSTM is shown in the Fig. (1).
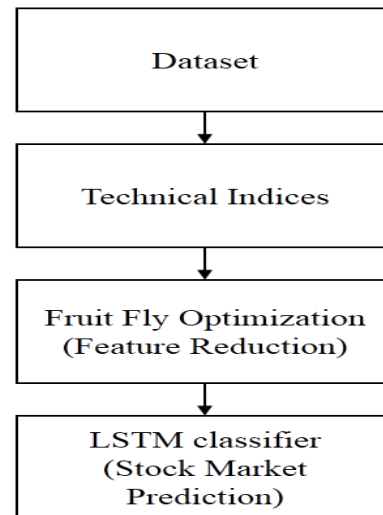

Fig. 1.The architecture of fruit fly optimization method with LSTM classifier

A. Technical Indices

The five technical indices of the stock market were used in this research for the stock market prediction. The five technical indicators are Exponential Moving Average (EMA), Average True Range (ATR), Commodity Channel Index (CCI), Price rate-of-change (ROC), and Relative Strength Index (RSI). Apart from these, the stock opening price and stock closing price were also used in this method for the prediction of stock market.

Technical Indices	Formula
Exponential Moving Average (EMA)	$EMA(N) = \begin{cases} S_{1,close} & \text{if } N=1 \\ \dfrac{2}{N+1}*S_N + \dfrac{N-1}{N+1}*EMA(N-1) & \text{if } N>1 \end{cases}$
Average True Range (ATR)	$ATR_t(n) = EMA_t(n,TR)$
Commodity Channel Index (CCI)	$(M_t - SM_t)/(0.015 \times D_t)$
Price rate-of-change (ROC)	C_t / C_{t-4}
Relative Strength Index (RSI)	$RSI(N) = 100 - 100/\left(1 + EMA(N)_{up}/\left(EMA(N)\right)\right)$ $EMA(N)_{up}$ is upward changes, $EMA(N)_{down}$ is downward changes for N

B. Fruit Fly Optimization

The Fruit Fly Optimization Algorithm (FOA) [14] is inspired by the natural activity of Drosophila foraging behavior. Fruit fly has superior olfactory capacity and visual sense, thus it's able to locate food by fully utilizing its instinct. The fruit fly nose is able to pick various food scents even from

a distance of 40 km. With the help of their sensitive visual organs, the fruit flies are able to locate food and the company's flocking location and then fly in that direction. The whole swarm system shares the best fruit fly's information during the iteration and the previous best fruit fly information is used for the next iteration.

The FOA can be categorized into several steps based on the fruit fly swarm food search process. This process is explained as follows:

Step 1: Initialization of Parameters

The parameters of FOA, such as population size, maximum iteration, the initial location (X_axis, Y_axis) and the random flight distance range are initialized, as shown in Eq. (1-2).

$$X\_axis = rands(1,2) \quad (1)$$

$$Y\_axis = rands(1,2) \quad (2)$$

Step 2: Initialization of Population

For individual fruit fly food search, the random location (X_i, Y_i) and distance is provided, where i denotes the size of population, as given in Eq. (3-4).

$$X_i = X\_axis + RandomValue \quad (3)$$

$$c_t \ Y_i = Y\_axis + Randomvalues \quad (4)$$

Step 3: Population evaluation

The distance of the food location to the origin (D) is calculated. The reciprocal of such distance is the smell concentration judgment value (S), as shown in Eq. (5-6).

$$D_i = \sqrt{X_i^2 + Y_i^2} \quad (5)$$

$$S_i = \frac{1}{D_i} \quad (6)$$

Step 4: Replacement

The S value is denoted in the smell concentration judgment function (fitness function), to find the individual fruit fly smell concentration $(smell_i)$.

$$Smell_i = Function(S_i) \quad (7)$$

Step 5: Identify the maximal smell concentration

Maximal smell concentration of fruit fly and corresponding fruit fly location is measured using the Eq. (8).

$$[bestSmellbestIndex] = \max(Smell) \quad (8)$$

Step 6: Keep the maximal smell concentration

The maximum smell concentration value and co-ordinates x and y are retained. Then, the fruit fly swarm flies towards the location of maximal smell concentration value, as in Eq. (9-11).

$$Smellbest = bestSmell \quad (9)$$

$$X\_axis = X(bestIndex) \quad (10)$$

$$Y\_axis = Y(bestIndex) \quad (11)$$

Step 7: Iterative optimization

The step 2-5 is continuously implemented until the maximum number of iteration. The process stops when fitness values is not higher than previous value or reaches the maximum iterative number.

C. Long-Short Term Memory

Stock market prediction method is not only based on recent data but is also required in the previous data. The RNN has advantages in handling the data with long-term dependencies, but there are some limitations in the practical applications. This is the benefit of the hidden layer in the self-feedback mechanism. LSTM consists of three gate (input, output and forget gate), which determines the data that need to be present in the memory cell [15].

The LSTM cell input data is denoted as x_t at time t and the LSTM cell output at the previous moment is denoted as h_{t-1}. The memory cell value is denoted as c_t and the LSTM cell output is denoted as h_t. The LSTM unit calculation process is categorized in the following steps.

1. The weight matrix W_c, the candidate memory cell c_t, and the bias b_c is calculated, as shown in Eq. (12).

$$c_t = \tanh(W_c.[h_{t-1}, x_t] + b_c) \quad (12)$$

2. The input gate i_t is calculated that update the input data to the memory cell state value. The weight matrix is denoted as W_i and the σ is sigmoid function, as shown in Eq. (13).

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (13)$$

3. The value of forget gate f_t is calculated, which controls the historical data update to the memory cell state value. Here, the bias is denoted as b_f and the weight matrix is denoted as W_f, as shown in Eq. (14).

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (14)$$

4. The memory cell current movement c_t and the last LSTM unit state value c_{t-1} is calculated, as shown in Eq. (15).

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (15)$$

Where ''*'' denotes the dot product. The input and forget gate control the memory cell update based on the candidate cell and previous cell state value.

5. The output gate o_t value is calculated that controls the output of the memory cell state value, as shown in Eq. (16). Here, the bias is denoted as b_o and the weight matrix is denoted as W_o.

$$o_t = \sigma(W_0.[h_{t-1}, x_t] + b_0) \quad (16)$$

6. The LSTM unit h_t output value is calculated in the Eq. (17).

$$h_t = o_t * \tanh(c_t) \quad (17)$$

The LSTM method uses three controlled gate to process read, reset and update long time information. Due to the LSTM sharing mechanism of internal process, the weight matrix is used to control the output.

Between input and feedback, LSTM applies a long time delay. The gradient will neither be reduced nor be removed due to the maintains of continuous error flow.

IV. EXPERIMENTAL DESIGN

Many researches have been carried out to increase the performance of stock market forecasting using machine learning methods. This study uses the feature reduction method based on fruit fly optimization. The LSTM is used for the stock market forecasting indices using the selected features of Fruit fly method. The experimental settings for evaluating the proposed framework are given in this section.

Dataset: The Nifty 50 and S&P 500 data were used to estimate the efficiency of the Firefly method based framework. The three year's stock data of Nifty 50 and S&P 500 from 01/11/2017 to 01/11/2019 were collected for estimation of model. The stock data were collected from the Yahoo website for the estimate of the firefly based method.

Metrics: The metrics such as Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were used to evaluate the model.

System Requirement: The proposed framework was analyzed in the PC consisting the processor of Intel i7 with 500 GB hard disk and 8 GB RAM. The model was developed using the python language, and Keras library has been used for LSTM.

V. EXPERIMENTAL RESULTS

Stock Market prediction method helps the investors in decision-making process to increase profit. Many existing method has been conducted on the stock market prediction and recently, deep learning method shows the significant performance. The fruit fly method has been proposed for the feature reduction process in the stock market forecasting and LSTM classifier is used for the prediction. Technical indicators were measured from the stock value and given as input to the developed framework. This section provides an analysis of performance of the developed model in the stock market prediction.
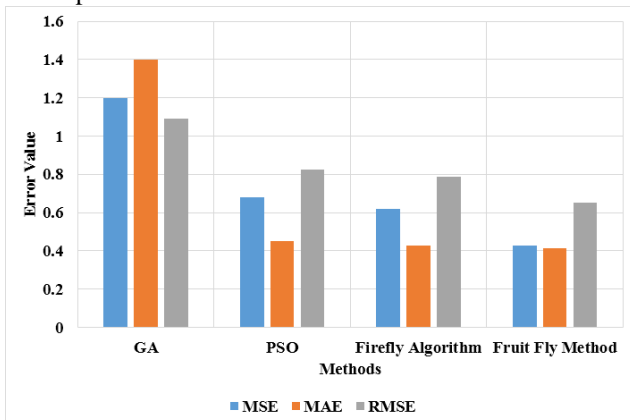

Fig. 2.The error value of fruit fly method in Nifty 50 data

The Nifty 50 data has been used to evaluate the efficiency of the framework and other existing optimization method, as shown in Fig. (2). The Fruit Fly based model has more effectiveness compared to other feature reduction method. The Firefly based method has achieved the second higher

efficiency in the stock market prediction. The three metrics such as MSE, MAE and MAPE were used to estimate the efficiency of the framework. The proposed fruit fly framework has MSE value of 0.426, while firefly method has a MSE value of 0.621. The fruit fly method has the significant performance than the firefly algorithm. The fruit fly method helps to overcome the overfitting problem in the LSTM method.

Table- I: The proposed method analysis in Nifty 50 data

Methods	MAPE (%)
GA	99
PSO	99
Firefly Algorithm	97
Fruit Fly Method	96

The proposed fruit fly method is evaluated in the Nifty 50 data with MAPE value, as shown in Table 1. The proposed fruit fly method has the lowest MAPE value compared to the other feature reduction method. The firefly method has the second higher efficiency in the stock market prediction. The fruit fly method has 96 % of MAPE, while the firefly method has 97 % of MAPE. The fruit fly method has the advantages of simple computation process and only less number of parameter is used for tuning. The fruit fly method helps to overcome the overfitting problem in the stock market prediction.
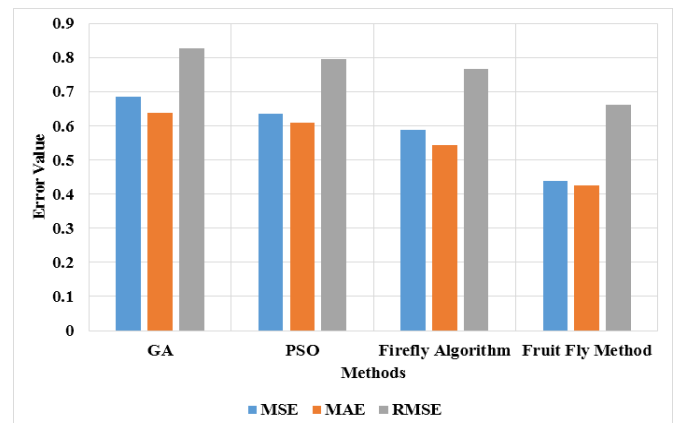

Fig. 3. The proposed Fruit Fly method in the S&P 500 data

Error metrics were evaluated from the proposed fruit fly method in the S&P 500 data and is compared in figure 3. The feature reduction methods such as GA, PSO, and Firefly is used for the comparison. The fruit fly method has the higher efficiency in stock market forecasting than other feature reduction method. The fruit fly method has the advantages of simple computation processes and less number of parameter for tuning. The fruit fly method selects the relevant feature for the stock market prediction and helps to reduce the overfitting problem in the LSTM classifier.

Table- II: The MAPE value in S&P 500 data

Methods	MAPE (%)
GA	99
PSO	99
Firefly Algorithm	98
Fruit Fly Method	96

960

The MAPE value is calculated for the proposed fruit fly method in S&P 500 data, as given in Table 2. The fruit fly method has a lower error value in the stock market prediction than other compared method. The fruit fly method has the advantages of simple computation process and less number of tuning parameters. The fruit fly method selects the relevant feature for the LSTM model in the forecasting.

Two stock data were used to evaluate the efficiency of the fruit fly method and tested with other optimization techniques. The evaluation shows that the fruit fly framework has more performance than existing methods.

VI. CONCLUSION

Stock market forecasting is one of the essential financial process that has gained researcher's attention for many years. The fruit fly method is proposed for the feature reduction in stock market prediction. The fruit fly method selects the relevant features for the LSTM to reduce the overfitting problem. The Nifty 50 and S&P 500 stock data were used to evaluate the proposed model efficiency. The technical indicators were used to analyze the stock value in a statistical manner. The experiment results show that the fruit fly method has more performance than other feature selection methods. The fruit fly method has MSE of 0.437 and the firefly method has 0.587 of MSE. In future work, sentimental analysis can be included to improve the performance of the developed model.

REFERENCES

1. S. K. Chandar, "Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction," *Journal of Ambient Intelligence and Humanized Computing,* 2019, pp. 1-9.
2. X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing*, 2018, pp. 1-21.
3. S. K. Chandar, "Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach," *Cluster Computing,* 2019, vol. 22, pp. 13159-13166.
4. I. Marković, M. Stojanović, J. Stanković, and M. Stanković, "Stock market trend prediction using AHP and weighted kernel LS-SVM," *Soft Computing,* 2017, vol. 21, pp. 5387-5398.
5. H. Chung, and K. S. Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," Neural Computing and Applications, pp.1-18.
6. Y. Song, J. W. Lee, and J. Lee, "A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction," *Applied Intelligence*, vol. 49, 2019, pp.897-911.
7. M. Jiang, J. Liu, L. Zhang, and C. Liu, "An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms," Physica A: Statistical Mechanics and its Applications, vol. 541, 2020, pp. 122272.
8. E. Chong, C. Han, and F.C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies", *Expert Systems with Applications,* vol. 83, 2017, pp.187-205.
9. Y. Baek, and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Systems with Applications*, 2018.113, pp. 457-480.
10. Y. Chen, and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction." *Expert Systems with Applications,* vol. 80, 2017, pp. 340-355.
11. S. R. Das, D. Mishra, and M. Rout, "Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method," *Expert Systems with Applications*: vol. 4, 2019, pp.100016.
12. R. Ramezanian, A. Peymanfar, and S. B. Ebrahimi, "An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market," *Applied Soft Computing*, vol. 82, 2019, pp. 105551.
13. W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowledge-Based Systems,* vol. 164, 2019, pp.163-173.
14. L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, B. Yang, and D. Liu, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, 2016, pp. 61-75.
15. J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on CEEMDAN and LSTM," *Physica A: Statistical Mechanics and its Applications,* vol. 519, 2019, pp.127-139.

AUTHORS PROFILE

Parag Rastogi is currently working as Assistant Professor in the Department of Computer Science and Engineering, SITE, Swami Vivekanand Subharti University, Meerut, UP, India. He has done M.Tech (CS), M.Sc. in Computer Science and B.Tech (CSE). He is a life member of ISTE, CSTA and IAENG. His areas of interest include Artificial Intelligence, Software Testing, Machine Learning and Cloud Computing.