An Improved Stock Prediction System Using Optimal parallel Cat Swarm Optimization Approach in Online Social Media Posts for Medicine Applications

Anitha J.[1*] and Kalaiarasu M.[2]
1. Department of IT, Sri Ramakrishna Engineering College, Coimbatore, INDIA
*****anithasri.srec@gmail.com

Abstract
In the business area, it has always been a challenging task to predict the exact daily price of the medicines and how it affects the stock market index for hospitals .So there is a great deal of research being conducted regarding the prediction of the direction of stock price index for medicine rates. In the stock market the stock index or stock market is referred to the measurement of the assessment of the medicine rates and usage of medicine. It is calculated from the cost of the particular medicine stocks.

To portray the market, the economic managers or the investors selected this tool for differentiating the return on particular speculation. To predict the stock value of medicine there are more number of research work have been conducted by the various researchers. Most of the approaches lacks from accurate prediction of stock measure value. In the proposed research work, efficient and accurate prediction of stock indices values of medicines are done by constructing the novel framework that can calculate and select the most optimal price rate of the medicines.

The performance accuracy of the proposed research work is improved by considering the social media posts related to medicine that can provide the most recent updates of the stock market levels. The parallel cat swarm optimization approach is utilized in this methodology for stock index value prediction of medicine.

The features of medicine that are considered for accurate prediction of stock index measurement values are, "closing price, opening price, lowest price, highest price, and total volume of stocks" in daily activities. Before stock prediction, fuzzy c means clustering is applied on the social media retrieved medicine posts for the feature selection purpose which can lead to selection of only most required contents. The experiments tests were conducted in the MATLAB simulation environment from which it is proved that the proposed research work can be provide better result than the existing methodologies such as BFO, ABFO and CSO.

Keywords: Stock Index, Noun Phrase, Opening Price, Closing Price, Optimal Detection.

Introduction

In the stock market the stock index or stock market is referred to the measurement of the assessment of the segment. It is calculated from the cost of the particular stocks classically weighted average. To portray the market, the economic managers or the investors are selected this tool, and to differentiate the return on particular speculation. An index is a numerically created, so it does not directly invest in the stock market. But in the stock market the number of mutual funds and exchange-traded funds endeavor to follow an index and those funds that do not may be judged in opposition to those that do [1].

Stock market index can be differentiating in more number of ways. In the world wide stock market index including the big companies without consider for where the stock market can be traded such as MSCI World and S&P Global 100. In the national market the 'national' index is referred to as the performance of the stock market and by deputy, reflects customer response on the circumstances of its social. The very important commonly quoted market indices are the national indices unruffled of the stocks of big companies scheduled on nation's biggest stock exchanges like American S&P 500, British FTSE 100 and Japanese Nikkei 225.

Some other local company's indices like FTST Developed Asia Pacific Index or FTSE Developed Europe Index. By using the exchange details only may be the index are created like NYSE US 100 or NASDAQ-100 or groups of exchanges like Euronext 100 or OMX Nordic 40. In the United States Company, the stocks almost openly traded and it is represented as Wilshire 5000 Index including the entire U.S. Stocks traded on the New York Stock Exchange, NASDAQ and American Stock Exchange. It does not contain the address or limited partnership details. Russell Global Index initiating by the Russell Investment Group including the family indices.

Further, the dedicated indices are already available to following the performance of particular segments of the stock market. For example, the Wilshire US REIT which finds more than 80 American real estate investment trusts and the thirty six American firms are available in the Morgan Stanley Biotech Index in the biotechnology commerce. Some other kind of indices may track the size of the companies, type of administration and so on. One index is

initiated by Linux Weekly News tracks and the information like the sell products and services on the Linux operating surroundings.

In the open stock market the index future cost is repeatedly a better indicator but the signal employs only for a described time. At the opening stage the trading is classically unstable, which the financial records for a inconsistent amount of total trading capacity. If the commercial customers are weighs to buy a big share or sell the share in the multiple stocks, the stock market brunt can devastate whatsoever cost society the index futures point out. Commercial investors always observe the cost of stock, other than the largest orders they have to carry out, a smaller amount of significant the index future direction signal becomes [2].

The index value can be affected due to open the stock market not on time. Even though the every stock market starts to trade at once, the market opens at 9:30 am. According to the sale procedure the opening cost is set to the stocks, the benefits do not have the common characteristics, the extra stock will be closed in anticipation of the identical orders come in.

If the upcoming cost will be increased in the market, the customers want to sell the index that day to wait in anticipation of subsequent to the market opens previous to entering their order or set a higher cost limit. Buyers possibly will crave to seize off the index while the index futures can expect he inferior opening. Not anything is definite, on the other hand, most of the time the index future can calculate the opening market direction, but even the most excellent soothsayers are not for eternity correct. In this proposed work, the stock prediction is mainly concentrating on to improve the status of company owners to give the correct data about the future stock index value. By using the clustering methodology the exact calculation of stock index value is preprocessed from the social media content namely FCM clustering method. After that by using this approach the stock value is predicted and it is known as parallel cat swarm optimization approach. This entire research work process is done in the mat lab simulation.

The research work is divided into the modules. The corresponding modules can be explained in the following subsection. The existing research work has been explained in the section 2. Present work is explained in detail in section 3. The research tests have been conducted in terms of various performance metrics in section 4. At lasts the conclusion of the entire research work in section 5.

Related Works

A lot of realistic time series data such as the commercial data are at the specified time incorrectly recorded and for this reason it treated as outliers. Observations are the outliers which are conspicuously vary from the extra data set. From the precedent data the outliers can be observed and this is used as training set may be as high as 40% [3]. Additionally at the same time as gathering and storing data, the position of and scale of altered form by some of the observations are also not recognized. Based on the position and scale, the observations may have responsible to harsh things on the performance of the prediction model.

Based on the learning techniques in the conventional adaptive forecasting models inclined such as the least mean square (LMS) [4] and back propagation (BP) [5] algorithms are used. These types of learning algorithms are inclined to the investigate type and it has been determined by using the square of the outcome fault as the price function. This result has been reported and then modernizes procedures determined from the easy price function are not automatic in opposition to the observations in the training data. Similarly, by using the degraded data sets the outcome model has been implemented and it exposes the bad prediction potential. In the occurrence of noisy training set the performance of the prediction can be increased and the price function is also very important. In the prose the Wilcoxon norm based regressor has been shown to be inducement to outliers [6].

In present days, using computing tools more number of researches has been published like Genetic Algorithm (GA), Bacterial Foraging Optimization (BFO), Genetic Programming (GP) and Particle Swarm Optimization (PSO) to implement the predicting methods. The stock data are extremely time-invariant and also extremely predisposed by undefined trade. To assortment optimization, economic failure calculation, pecuniary prediction, hoax revealing and scheduling purpose the more number of computation tools will be used. For the portfolio optimization the genetic algorithm is used to implement the prediction models [7]. In this current paper [8,12] an effective enquiry method in portfolio administration using Genetic Algorithm has been implemented. To calculate the exchange rates in the stock market the genetic programming based trading model is implemented by Bhattacharya et al [9].

To determine the cost formulas using exact information from S&P 500 index for training and testing by using genetic programming in some other employ by Chen et al. [10]. To achieve the implementation of the effective model to calculating the different stock indices by using Bacterial Foraging Optimization (BFO) and Particle Swarm Optimization (PSO) in [11] & [13] correspondingly.

Efficient and Accurate Stock Index Value Prediction System: In the marketable application, the economic forecasting or the particularly Stock Market Prediction is one of very popular area and it gives gorgeous reimbursement. In the Stock market lot of money is being invested and the customers get anxious and apprehensive of the upcoming trends of the stock prices in the market. The most important vicinity of anxiety is to decide the proper time to buy the share, hold or sell. Unluckily, stock market prediction is not a simple job because the stock market indices are basically dynamic, nonlinear, complex, nonparametric and

disorganized. In the commercial world, the stock index value prediction is the most important application which is used to focus the company's earnings stage.

In this research work, stock prediction is concentrated to improve the strategic improvement of the companies by constructing the novel framework. The proposed framework processes on the social media contents about the stock prices to improve the accuracy of the system. The optimization algorithm is used for the better prediction of the stock index value in terms of various features. However the stock prediction on the social contents would be more difficult process due to presence of various irrelevant features and data's that are present in the data set. This is resolved by introducing the filtering process which is done based on the noun phrases that are present in the social media data set. And then the remaining data values would be processed for the feature selection for retrieving only the contents that are related to the corresponding stock management system. Finally the parallel cat swarm optimization approach based stock value prediction would be done to accurately predict the stock index value, so that the company can attain more profit. The overall processing steps of the proposed research work are given in the following figure 1.

The figure 1 reveals the processing procedure that has been carried over to perform the stock prediction accurately from the social media contents that are posted by various researchers from the different part of the world. The social media contents would consists of different type of post which might include both relevant and irrelevant posts to the stock indices raises. Processing of these entire social media content would lead to the performance degradation in terms of more time consumption while processing the data's that are irrelevant to the stock related contents. The processing steps of the proposed research methodologies are given as follows:

- Gathering and storing the twitter data set about the stock index value discussion pages.
- Perform filtering on social media data set to remove the data's that are irrelevant to the stock pages using noun phrase concept.
- Perform feature selection on the data's that are resultant from the filtering process using Fuzzy C Means clustering approach.
- Predict the stock index value by training and testing the stock related posts using Parallel Cat Swarm Optimization approach.

The processing steps of stock index value prediction are given in the preceding points. From these points it is clear that the proposed research work can perform stock index prediction can be done accurately in the given proposed system. As given above, the first step is to gather and store the stock related contents from the online web social media related sites. This process would retrieve the post contents from the pages of stock market related person social media accounts. In this research work, social media website is considered for the retrieval and prediction of stock market index values. From the twitter accounts, the stock marketing user's pages would be analyzed and the posts submitted by them would be extracted. The remaining steps such as filtering feature selection and the stock index prediction is explained detailed in the following sub sections.

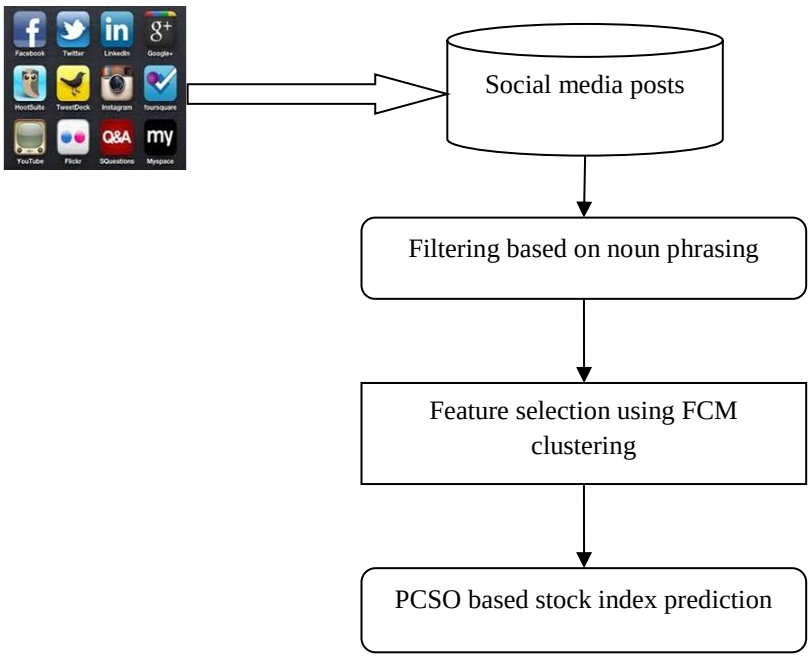

Figure 1: Overall Flow of the Proposed Stock Prediction System

Filtering of data using noun phrasing approach

An impartiality market or the share market and the stock market are the aggregation of buyers and sells of stocks. The stock is also known as shares. A stock exchange is a association by the companies or peoples called stock traders can trade stocks. Companies need a stock exchange details in the list manner.

Through the dealers other stocks may be traded "over the counter" (otc). Across the world, the big companies commonly have the stock exchange details.
In the twitter social media website, the related information in the stock market is extracted by using the filtering method. In this filtering method is used to decrease the computational transparency happen owing to processing not related data's. In this present research work, by using the different parameters the twitter is calculating the stock values. Twitter is one of the most popular social networking areas. By using this twitter application, the user can send and receive the message up to 140 characters and it is known as "tweets". To send and post the message process can applicable only the authorized users but in this case the unauthorized used can only permit to read the message.

There are 6,159,239 inimitable word stipulation are available in the Twitter data set. For example strings of non-symbolic typescript. These kinds of additional data require elevated computational price and also the threat of in excess of apt. subsequently; in the Twitter dataset the simple messages create the information alteration. Only the limited information is related to the stock market within the sixty one million messages and from this information the related stocks are selected for the prediction process. The noun phrasing is used to differentiate the related and unrelated messages. The noun phrase method is treated to the filtering step of this present work. In the Twitter post the nouns are recognized and this information is stored in the data set for supplementary processing only if the intention stock having the corporation name.

Based on the noun phrase to processing the filtering method in this present work the probabilistic method is developed. By using the LOB Corpus values the words are combined in this probabilistic methodology. The expansion of LOB is Lancaster-Olso/Bergen Corpus. In 1970s, a lot of words are collected for the British English texts and it was compiled. It is teamwork between the University of Olso, University of Lancaster and the Norwegian Computing Centre for the Humanities, Bergen. In 1960s, the Henry Kucera and W.Nelson Francis are compiled the British complement to the Brown Corpus for American English. The British Authors are premeditated to equivalent the real Brown Corpus based on its range and genres by using the documents available in the United Kingdom. After that based on the tag value the noun phrase will be extracted. Figure 2 demonstrate the filtering process.

STEPS: noun phrase extraction

1. Combine the input sentences.
2. By using a probabilistic partial parser the input sentences are divided into the chunks.
3. Each and every chunk makes a decision of syntactic head and semantic head.
4. From these chunks the noun phrase is extracted corresponding to the syntactic and semantic heads and these data's are linked to more number of noun phrases by a finite state mechanism.

Each and every word the input is applied in the POS and then linked into a chunk. The chunk will produce a output sequences. At last, a finite state NPTRACTOR will extract NPs.

Feature selection using FCM clustering approach

To predict the stock index value from the stock data set the features are extracted for the reason of to calculate the exact value of the stock prediction. This process can be achieved by combining the data based on the unique characteristics. For that reason the unwanted information are removed from the stock dataset. In hypothesis, after the filtering process the feature selection process is considered to second preprocessing step. To determine and categorize the structures is known as clusters, in the dataset by identifying and classifying the uniqueness between the separate data points namely patterns.

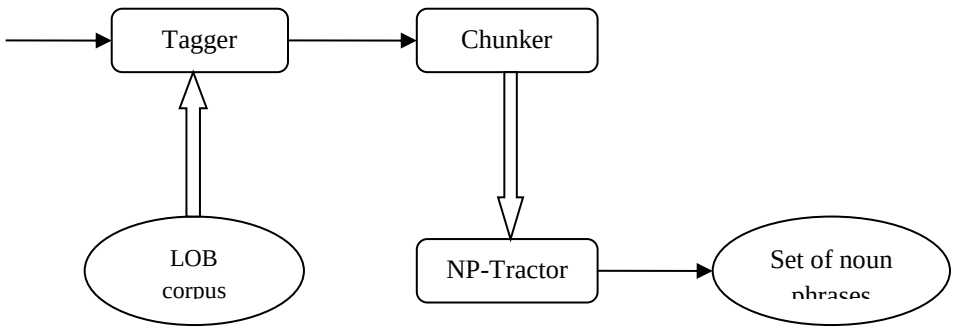

Figure 2: Processing Steps of Noun Phrase Extraction

The more number of clustering algorithm is used to divide the datasets. The general thought of the clustering algorithm is to identify prototype known as set of points in the region of which cluster is used to increase and to collect the pattern membership to clusters to get a good partition. Usually, this process is unsubstantiated. It means it happen without any intimation. The membership value are either 0 or 1 during the two-valued clustering process, in the fuzzy clustering process the exact value between the 0 & 1. FCM algorithm is one of the very popular algorithms of fuzzy clustering algorithm.

Consider, $X = \{x_1, \ldots, x_N\}$ is the dataset with $x_k = (x_{k,1}, \ldots, x_{k,F}) \in R^F$. consider A1,…..,AC is the same family members of fuzzy clusters on X and $V_{C,N}$ is represented as C X N matrices, where $C \in \{2..N-1\}$. The following statement is expressed in terms of the fuzzy C-partition of X elements in the data set

$$M_c = \left\{ U \in V_{C,N} \middle| u_{i,k} \in [0,1] \forall i,k; \sum_{i=1}^{C} u_{i,k} = 1 \ \forall k; 0 < \sum_{k=1}^{N} u_{i,k} < N \ \forall i \right\}$$

In the above equation the membership value of x_k to Ai is represented as $u_{i,k}$. consider the clusters A_i are spherical or ellipsoidal.

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, \quad 1 \le m < \infty$$

Any real number can be represented as m and it is greater than 1, the degree of association of x_i in the cluster j is represented as u_{ij} and i^{th} of d-dimensional of measured data is represented as x_i, the d-dimensional center of the cluster is represented as c_j and finally the denominator ||*|| is expressed the uniqueness between the every measured data and the centre. The above formula is used to Fuzzy partitioning and it is carried in the course of an iterative optimization with modernize of membership u_{ij} and the cluster c_j and it is followed by,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

If the ε is represented as execution criterion and this value between 0 and 1 means the iteration process will be stopped when$\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon$. While the iteration step is mentioned by k. this algorithm converges to a local minimum or the saddle point of J_m. the following steps are explained the Fuzzy C-Means algorithm. There are,

Steps: fuzzy c-means

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: estimate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$
   $$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$
3. modernize $U^{(k)}$, $U^{(k+1)}$
   $$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}$$
4. If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ then STOP; or else go back to step 2.

Stock index value prediction using PCSO: To calculating from the different types of stock indices the PCSO model is used to implement the prediction model in this proposed work. By minimizing the mean square error (MSE) the linking weights of the adaptive linear combiner based models are calculated by using PCSO. By using the adaptive linear combination method the novel methods are implemented to calculate the stock indices and the PCO. Figure 3 illustrate the basic structure for the stock market prediction and it is considered to be an adaptive linear combiner with parallel inputs.

Figure 3 illustrate the entire process of the stock prediction system based on the adaptive linear model. At first, the input features are provided into the network. In this current work, the input features participate a major role and it is used to calculate the stock index value consequent corporation in the exact method. There are five input features are used in this present work. They are,

- Opening price.
- Closing price.
- Highest price.
- Lowest price.
- Total volume of stocks.

From the Twitter dataset, these five types of input features are extracted from the particular stage and then these input features are given an input to the network. After this process, the weight value is predicted for this input features after that the sum of the weighted value is determined. This prediction process can be achieved by using the parallel cat swarm optimization methodology.

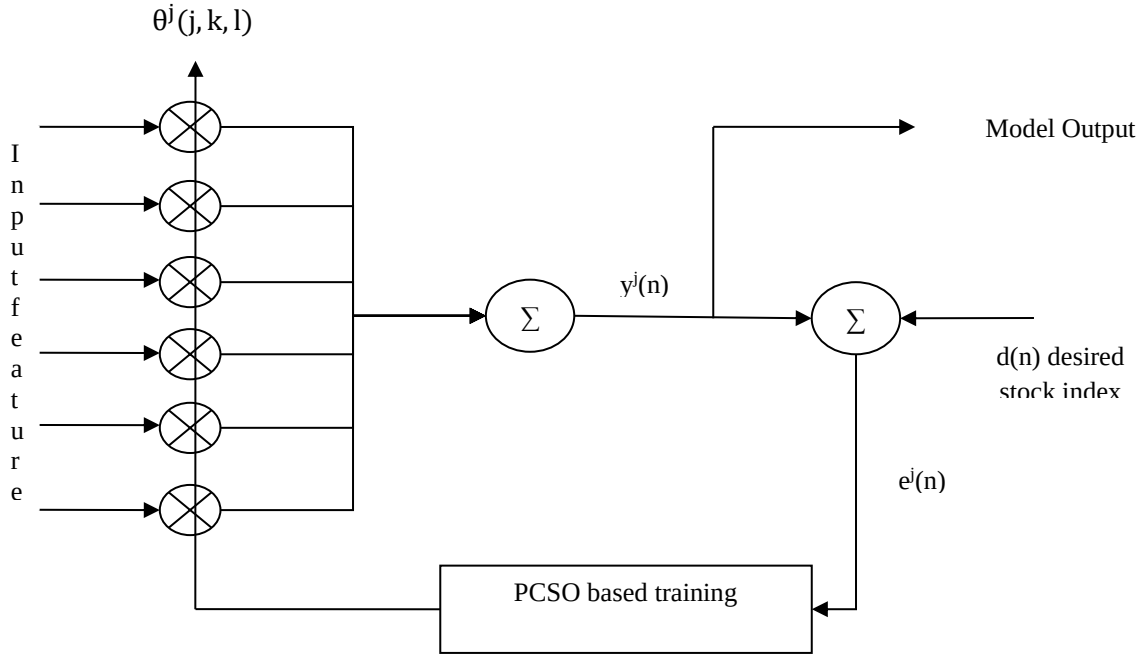

Figure 3: Adaptive Linear Model

Parallel cat swarm optimization approach

Based on the CAT algorithm the CSO algorithm is implemented. To resolve the optimization trouble, the CSO algorithm is used. Each and every cat positions composed of M dimension, velocities for every dimension and a fitness value on behalf of the space of the cat to the benchmark function and a flag to validate whether the cat is in seeking mode or tracing mode. The CSO algorithm maintains the every iteration steps until it get the better result. There are two sub modes are used to CSO algorithm. The first one is seeking mode and the second one is tracing mode.

By observing the model of CSO, the persons employ separately when the persons hang about in the hunt for manner. On the other hand, the unique data's are split between them corresponding to the tracking mode. Mostly the individuals employ a stand next to the system under the model of CSO. To parallelize the model for CSO then the algorithm of tracking mode can be alerted to create it seems supplementary supportive. At first, the individuals are alienated into the sub-populations in PCSO. For this reason, in the tracking mode the individuals do not shift frontward to the global best solution in a straight line, but they shift frontward to the local best solution of its have possession of collection in common.

Fundamentally, compared to the CSO and the PCSO the important structure is similar to both of them. At first, N number of individuals are created and divided into G groups. Consider, the G groups is equal to 1, after that the PCSO go back into the real CSO.

Seeking mode

There are four types of seeking mode is used in CSO such as Seeking Memory Pool (SMP), Counts of Dimension to Change (CDC), Self Position Consideration (SPC) and in search of range of the selected Dimension (SRD). Seeking memory pool is used to describe the size of seeking memory of every cat and it is also indicate the some other points minimized by the cat. From the memory pool the point can be selected by using the cat corresponding to the policy. To select the dimensions for the mutative ration process is declared by SRD. For the mutation process if the dimension has been selected means the dissimilar between the previous and future value range will comes under the specified range and this range value is described by SRD. CDC is used to tell the information about the number of various dimensions. The Boolean valued variable is known as SPC and it is used to check whether the point value is already standing of the person's points to move to. SPC cannot use the value of SMP. The following steps are described the seeking mode:

1. from the position of cat_k to create the j copies, where j = SMP. If the value of SPC is true, let j = (SMP-1), then keep the current position as one of the candidates.
2. For every copy, corresponding to CDC, arbitrarily plus point or negative SRD percents the current values and substitute the older ones.
3. Estimate the fitness values (FS) of every candidate points.
4. Predict the probability of every candidate point by using equation (1) if the FS value is not equal or else to set the probability value to the candidate point is 1.
5. Altered the position of cat_k and consequently select the candidate points.

$P_i = \frac{|FS_i - FS_b|}{FS_{max} - FS_{min}}$, where 0<i<j

The aim of the fitness function is to identify the smallest amount of result, consider $FS_b = FS_{max}$, or else $FS_b = FS_{min}$.

Parallel tracing mode process

The following steps are explains the parallel tracing mode process. The steps are,

1. Corresponding to the equation (4) at the present iteration process update the velocities for each and every dimension $v_{k,d}$ (t) for the cat_k
2. Validate the velocities values are having the highest velocity. If the values is extremely high means than the values are set equal to the limit.
3. Corresponding to equation (5), the position of cat_k is updated.

$$V_{k,d}(t) = V_{k,d}(t-1) + r_1.c_1.\left[x_{lbest,d}(t-1) - x_{k,d}(t-1)\right], d = 1,2,\dots M$$

The position of the cat is represented by $x_{lbest,d}(t-1)$ and who has the best fitness value, at the past iteration in the group that cat_k belongs to.

$$x_{k,d}(t) = x_{k,d}(t-1) + v_{k,d}(t)$$

Based on the different indicator values the stock index value can be predicted accurately. The process of to calculate the stock index value is explained in the above algorithm. The objective function is treated as a indicator in the stock market surroundings and it is used to give the simple surroundings for the corporation to calculate the altered stock values. The following procedures explain the objective function.

Exponential moving average = (P× A) + (Previous EMA × (1-A));

A = 2 / (N+1)

Where P-Current price, A-Smoothing factor, N-time period

Accumulation/distribution oscillator = $\frac{(CP-LP)-(HP-CP)}{(HP-LP)\times(Period's volume)}$

Where CP-Closing price, HP-Highest price, LP-Lowest price

Stochastic indicator (STI) = %K = $\frac{(Today's close - Lowest\ low\ in\ K\ period)}{((Highest\ high\ in\ K\ period - Lowest\ low\ in\ k\ period))} \times 100$

Relative strength index RSI = $100 - \frac{100}{1+(U+D)}$

Where U = total gain/n, D = total losses/n, n= number of RSI period

Price rate of change (PROC) = $\frac{(Today's close - close\ x - period\ ago)}{(closrX - period\ ago)} \times 100$

Closing price acceleration = $\frac{(close\ price - close\ price\ N - period\ ago)}{(close\ price\ N - period\ ago)} \times 100$

High price acceleration = $\frac{(High\ price - High\ price\ N - period\ ago)}{(High\ price\ N - period\ ago)} \times 100$

Experimental Results

To analyze the performance improvement of the present and previous research methods in terms of the different performance measures by using the arithmetical calculation. The future research work can be implemented in the mat lab simulation background. The stock marketing related details are available in the twitter dataset and this information is extracted from the different places for to get the best prediction of the stock index value. The performance measures painstaking in this work are scheduled as follows:

- Accuracy.
- Jaccard Index.
- Rand index.
- Root mean square error.

The above mentioned performance are estimated for both the present methods namely PCSO and the previous methods namely CSO, ABFO, BFO. These methods are illustrated in the graphical representation from figure 4 to figure 8. These measures obviously point to that the PCSO based model offers best ever convergence at the same time as training followed by CSO, ABFO and then BFO models. Supplementary, corresponding to the MSE method gain the models are integrated as PCSO/CSE, ABFO and BFO. Compared to these models the PCSO give a best result.

The table 1 & 2 explains the arithmetical measure value that is calculated from the estimation of present and previous methods in terms of the different performance measures.

Table 1
Accuracy, Jaccard Index Comparison Values

Data Size in MB	Accuracy values				Jaccard Index values			
	PCSO	CSO	ABFO	BFO	PCSO	CSO	ABFO	BFO
100	0.86	0.78	0.71	0.62	0.7125	0.6826	0.6723	0.6099
200	0.82	0.81	0.72	0.60	0.4659	0.4258	0.3950	0.3870
300	0.74	0.71	0.69	0.65	0.6874	0..5670	0.5326	0.5240
400	0.72	0.65	0.61	0.52	0.5269	0.4379	0.4141	0.3959
500	0.69	0.63	0.59	0.54	0.8253	0.6972	0.5705	0.5505

Table 2
Rand index, and Root Mean square value comparison values

Data Size in MB	Rand Index values				RMSE values			
	PCSO	CSO	ABFO	BFO	PCSO	CSO	ABFO	BFO
100	0.6925	0.6846	0.6723	0.6171	0.36	0.39	0.45	0.56
200	0.5659	0.5321	0.4998	0.4822	0.42	0.45	0.49	0.51
300	0.7874	0.7278	0.6971	0.6853	0.49	0.51	0.51	0.55
400	0.6279	0.5962	0.5312	0.5144	0.52	0.56	0.61	0.66
500	0.8653	0.7892	0.7429	0.7029	0.53	0.56	0.59	0.69

Accuracy comparison

From the various stage of time the exact value is described as the right prediction of the stock index value from the different stock index values. Compared to the previous methods like CSO, ABFO and BFO the exact value can be predicted in this present system. By using the stock index value prediction system, the exact value is determined true positive, false positive, true negative and false negative values. The following formula defined to calculate the accuracy.

$$\text{Accuracy} = \frac{T_p}{(T_p + F_p + F_n)}$$

The graphical illustration of accuracy and data size is given as like as follows:

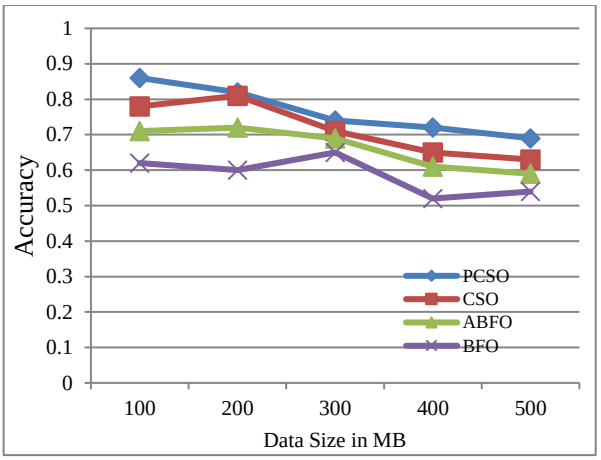

Figure 4. Accuracy Vs data size

Figure 4 explains that the accuracy of the present approach specifically PCSO is superior to the previous approach specifically CSO, ABFO and BFO. Compared to the previous approaches the present approach have the exact value for the various set of data sizes.

Jaccard index comparison

To characterize the relationship between the data points the jaccard index is used. The jaccard index value is used to calculate the exact stock index value. The jaccard coefficient estimates the relationship between the restricted sample set, and it is described as the size of the intersection separated by the size of the union of the sample set:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

In the above equation A, B represents the data points.
In the present work, the jaccard index value wants to be improved greater than before, which indicates that the present approach performed stock index value calculation with increased exact value.

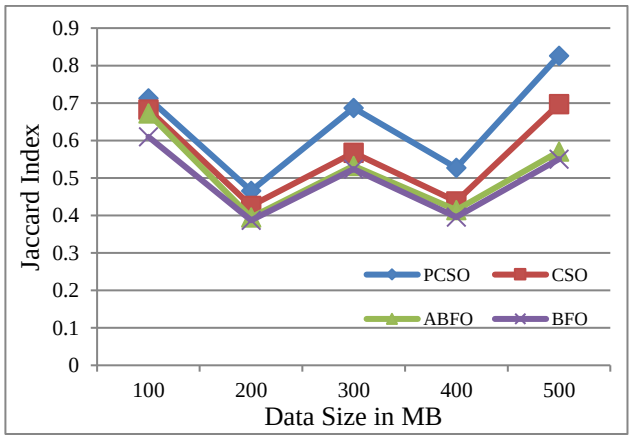

Figure. 5 Jaccard Index Vs Data Size

Figure 5 illustrate the Jaccard Index Vs data size. It explains that the stock prediction is completed in exact solution in the proposed methodology that is PCSO compared to the present

methods like CSO, ABFO and BFO. The Jaccard index is greater than before in the proposed method simply than the existing methodology for the various set of data sizes.

Rand index comparison

To find the uniqueness between the two predicted values by using the Rand Index or Rand measure while the jaccard index calculating the exact result of the single prediction value. {\displaystyle R={\frac {a+b}{a+b+c+d}}={\frac {a+b}{n \choose 2}}}Given a set of n elements S = {o_1, o_2, …, o_n} and two partitions of S to compare, X = {X_1,…, X_r}, a partition of S into r sub sets, and Y = {Y_1,…, Y_s}, a partitions of S into s sub sets, define the following:

1. The number of pairs of elements in S that are in the same set in X and in the same set in Y.
2. The number of pairs of elements in S that are in the different set in X and in the different set in Y.
3. The number of pairs of elements in S that are in the same set in X and in the different set in Y.
4. The number of pairs of elements in S that are in the different set in X and in the same set in Y.

The rand index R is computed as like follows:

$$R = \frac{a+b}{\binom{n}{2}}$$

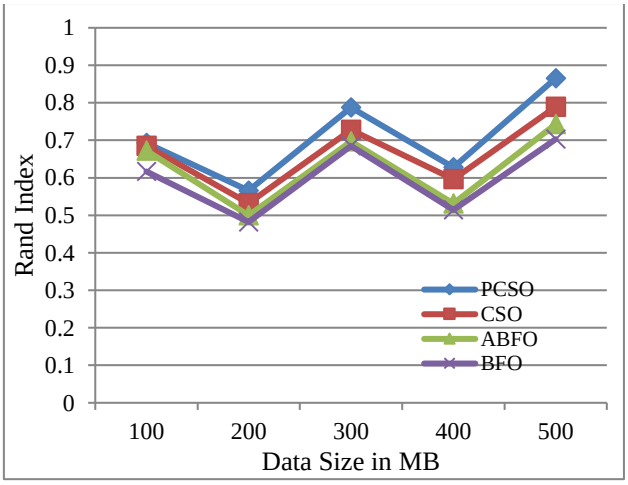

Figure 6: Rand Index Vs Data Size

Figure 6 illustrate the Rand Index Vs data size. It explains that the stock prediction is completed in exact solution in the proposed methodology that is PCSO compared to the present methods like CSO, ABFO and BFO. The rand index greater than before in the proposed method simply than the existing methodology for the various set of data sizes.

Root mean square error comparison

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a recurrently used to compute the dissimilar between the predicted values by a reproduction or an estimator. The RMSD represents the sample standard deviation of the dissimilar between the calculated values and focused values. This unique difference is known as residuals when the predictions are performed in excess of the data illustration that was used for the estimation. The RMSD treated to aggregate the prediction errors for different times into a single measure of predictive supremacy. To measure the exact result the RMSD is the best one but it is only to differentiate the prediction fault of various models for a specified variable and not between the variables, as it is scale – dependent.

The RMSD of predicted values $\hat{y}_t$ for times t of a regression's dependent variable y_t is computed for n different predictions as the square root of the mean of the squares of the deviations:

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$$

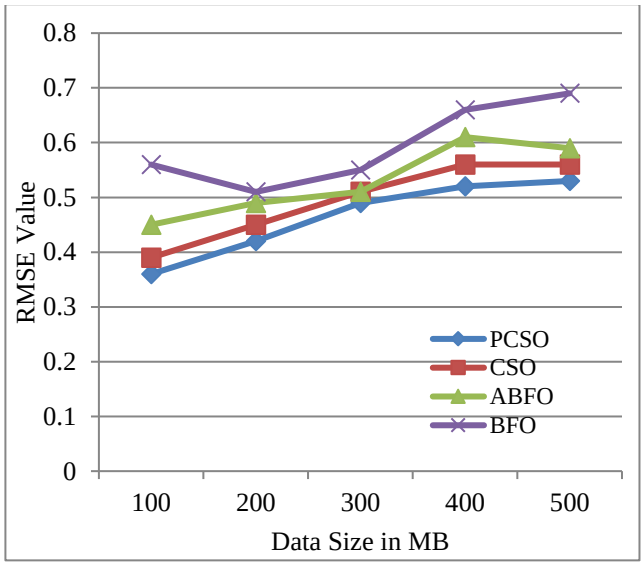

Figure 7: RMSE Vs Data Size

Figure 7 illustrate the RMSE Vs data size and it can be proved that root mean square fault value of the future method is significantly lesser than the presented methods namely CSO, ABFO and BFO.

Conclusion

Public health agencies conduct stock market analysis to detect prediction rate. Typically, the collected information is used to help public health officials, researchers, and health care professionals cope with and plan for future stock market prediction. An impartiality market, share market or the stock market is the aggregation of buyers and sellers of stocks. At the current situation in the real world the stock index prediction is the very difficult task because it has the more number of applications turnover level depends on the exact calculations of the stock index measure values. In the future work, the parallel cat swarm optimization methodology is used to guide to good and exact prediction of the stock index value in the stock prediction structure. This proposed research work, the information is used by the social media to get the up to date data's about the increase/decrease stock

index measure values. The twitter information's are processed for the exact stock index prediction which is preprocessed to get the only the applicable information's to eliminate the computational operating cost. To retrieve the related information the noun phrasing and the feature selection is used. Compared to the other existing methodology, this proposed work guides to give the good result in terms of greater than before correctness and condensed fault value.

Reference

1. https://en.wikipedia.org/wiki/Stock_market_index.

2. http://www.investopedia.com/articles/active trading/070113/using-index-futures-predict-future.asp.

3. Hsieh, J. G., Lin, Y. L., & Jeng, J. H. (2008). Preliminary study on Wilcoxon learning machines. IEEE transactions on Neural Networks, 19(2), 201-211.

4. Widrow, B., & Stearns, S. D. (1985). Adaptive signal processing (Vol. 15). Englewood Cliffs, NJ: Prentice-hall.

5. Tan, C. N., & Wittig, G. E. (1993). A study of the parameters of a back propagation stock price prediction model. In First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, pp. 288-291.

6. McKean, J. W. (2004). Robust analysis of linear models. Statistical Science, 19(4), 562-570.

7. Lettau, M. (1997). Explaining the facts with adaptive agents: The case of mutual fund flows. Journal of Economic Dynamics and Control, 21(7), 1117-1147.

8. Jiang, R., & Szeto, K. Y. (2002). Discovering investment strategies in portfolio management: a genetic algorithm approach. In 9th International Conference on Neural Information Processing, 3, pp. 1206-1210.

9. Bhattacharyya, S., Pictet, O., & Zumbach, G. (1998). Representational semantics for genetic programming based learning in high-frequency financial data. In 3rd Annual Conf on Genetic Programming pp. 11-16.

10. Chen, S. H., Yeh, C. H. &. Lee, W. C. (1998). Operation pricing with genetic Programming. In Conference on 3rd Annual Genetic Programming.

11. Majhi, R., Panda, G., Sahoo, G., Dash, P. K., & Das, D. P. (2007). Stock market prediction of S&P 500 and DJIA using bacterial foraging optimization technique. In IEEE Congress on Evolutionary Computation, pp. 2569-2575.

12. Rajan, C & Shanthi, N, (2015). Genetic based Optimization for multicast Routing algorithm for Manet' Sadhana - Academy Proceedings in Engineering Science, Volume 40 Issue 7, 2341-2352.

13. Majhi, R., Panda, G., Sahoo, G., & Panda, A. (2008). On the development of improved adaptive models for efficient prediction of stock indices using clonal-PSO (CPSO) and PSO techniques. International Journal of Business Forecasting and Marketing Intelligence, 1(1), 50-67.

14. Rajan, C, Geetha, K, (2016), An Analysis of Compression Methods and implementation of Medical Images in Wireless Network", World Academy of Science, Engineering and Technology, Vol. 9, no. 9, pp. 2127-2133