

# PSO Swarm intelligence technique to optimized ANN for demand forecasting

Daham Owaid Matrood<sup>1</sup>, Muna Jaffer Sedeeq<sup>2</sup>  
[daham.stat@gmail.com](mailto:daham.stat@gmail.com) , [muna.jaffer@uomosul.edu.iq](mailto:muna.jaffer@uomosul.edu.iq)

Technical College of Management /Mosul, Northern Technical University, Iraq<sup>1</sup>, Computer Science Department, College of Computer and Mathematics Science, University of Mosul, Mosul/Iraq<sup>2</sup>

**Abstract.** With the expansion and development of artificial intelligence algorithms and the use of swarm intelligence and artificial neural networks extensively in solving many complex problems recently. In this research, one of the swarm intelligence algorithms was used, which is represented by the algorithm of the Particle swarm optimization (PSO), and then combined it with one of the algorithms of artificial neural networks, that represented by the algorithm of error back propagation neural network EBPNN, in order to solve the problem of forecasting the demand. And the data that was used in this research was prepared by the general company for prepared clothes and the northern general company for cement represented by the weekly demand data for cement and towels. And the method of the combined PSO with EBPNN obtained the better performance than the standard back propagation neural network algorithm.

**Keywords:** Swarm intelligence, optimization, Artificial Neural Network, Backpropagation

## I. Introduction

Realizing future better has pulled many folks for thousands of years The prognostication strategies vary greatly and will rely upon the availability of the information, the nature of models accessible, the types of presumptions offered, typically speaking prognostication isn't a simple task and it's attracted several researchers to investigate it.

Artificial neural network (ANN) has discovered expanding attention in prediction theory, resulting in effective applications in numerous foretelling domains together with economic, business money [1, 12]. ANN will learn from models and use them to predict future qualities. Additionally to it, they are able to handle deficient data and can be powerful particularly in circumstances wherever it's inconceivable to determine the principles or steps that cause the solution of trouble.

Back Propagation (BP) is one of the best techniques to coach ANN. A typical technique for preparing a neural network where in the underlying framework yield is compared with the ideal yield and the structure is balanced until the distinction between the two is restricted. In this way, BP ANN has exceptionally solid abilities of nonlinear demonstrating and examination for immense and complex framework [13].

However, since the underlying interconnecting loads of ANN are frequently stochastically given, the learning times and last interconnecting loads of the system are in this manner changed for various occasions of preparing. In other words, the prepared system isn't interesting and once

in a while the system conceivably dives into fractional least. Furthermore, the visual deficiency of the assurance of starting interconnecting loads consistently brings about too many preparing times and moderate assembly. These deficiencies of ANN truly sway its accuracy of demonstrating and impacts of use. It is very important to enhance and improve the loads of ANN [9].

In this paper, a particle swarm optimization (PSO) has been used to optimize the weights of ANN for demand forecasting, and exhibits an incorporated model for both ANN and PSO. As a case study, this model is applied to predict forecast weekly demand from packed cement and towels, which has been outfitted by the Northern General Company for Cement and General Company of prepared clothes respectively.

The paper is organized as follows: Section 2 briefly introduces the basic foundations of ANN, and the particle swarm optimization for training ANN has been proffered in Section 3. In Section 4, Experiments and discussion have been created for an industrial case study. The Comparison between HPSO and GA is given in Section 5. The conclusion of this study has been presented in Section 6.

## II. ARTIFICIAL NEURAL NETWORKS

Artificial neural systems are unquestionably one of the greatest powerful research territories during the cutting edge time frame, which has pulled in the consideration of numerous experts of various scientific fields. Ongoing examination exercises with respect to the artificial neural systems demonstrated that this strategy is an exceptionally amazing gadget to illuminate muddled issues viable in all engineering fields [10]. Neural system, like human mind, has the necessary capacity for learning and can use the procured new encounters from new and comparative issues. Despite the fact that ANNs are not similar to the human brain, these systems prepared unique highlights in some applications abilities, for example, division of pattern and its agreeability to gain proficiency with the systems by direct and nonlinear mapping any place the learning is required. Among the neural systems highlights and capacities, it tends to be alluded to certain cases including its agreeability of learning and its adaptability to accessible data, enormously equal handling the system inputs, therefore to quicken preparing time, high problem resistance, etc [11].

The design of the entire system changes frequently and adjusts to changing or new levels of data. furthermore, this system is enormously equal, powerful, fault tolerant and amiable to learning

Artificial neural networks could be a framework outfitted with a parallel processor for amassing mass-enormous information and that they work as a human cerebrum. The following represent the ANNs essentials:

- information are prepared in the solitary units named "node".
- The signals are moved between nodes via association lines.
- The weight ascribed to every association lines showed the communication capacity of that line.
- every node generally has initiation and move capacities, so as to determine yield signals from input information of the system [12].

The configuration of artificial neural networks is presented by association patterns between nodes, a strategy deciding the association loads and activation function. The structure of the artificial neural system has been generally framed by input layers, middle layers and yield layers (Fig. 1). Input layer is a transfer layer for giving information. The last layer (yield layer) incorporates the anticipated values by the system. middle layers include some processing nodes, where information is handled [11].

In this paper, we've utilized neural system with one hidden layer, the specification of number of input nodes, which are slacked perceptions of a similar variable; assume an urgent job as it helps in demonstrating the autocorrelation structure of the information. The specification of the number of yield nodes is moderately simple, during this paper, one yield node has been utilized and multi-step ahead prediction has been utilizing the repetitive technique as utilized in Box-Jenkins strategy. This includes utilization of expected value as an input for predicting the future value. It is in every case better to choose the model with a tiny number of nodes at concealed layer as it enhances the out of the sample prognostication achievement and furthermore maintains a strategic from the issue of overfitting [6].

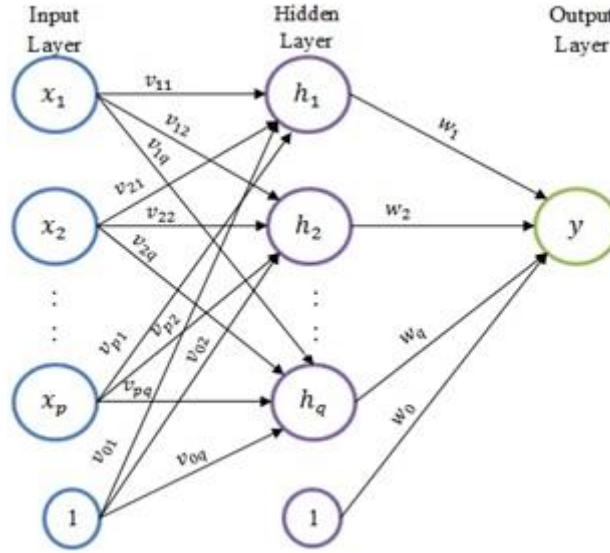


Fig. 1. Structure of neural network.

The general expression for the final output value  $y$  in a multi-layer feed-forward neural network is given by

$$y = g \left( \sum_{j=0}^q w_j f \left( \sum_{i=0}^p v_{ij} x_i \right) \right) \quad (1)$$

Where  $f$  and  $g$  indicate activation function at hidden and yield layer severally,  $p$  is a number of input nodes,  $q$  is the number of concealed nodes, ( $v_{ij}$  is that the weight connected to the association between  $i$ th input node to the  $j$ th node of concealed layer,  $w_j$  is the weight joined to the association from  $j$ th concealed node to the yield node and  $x_i$  is the  $i$ th input of the model.

Every node of the hidden layer gets the whole weighted of all inputs including an inclination term for which the estimation of input variable will consistently take a value one. This whole weighted of information is converted by each concealed node using  $f$ . In a very similar manner, the yield node gets the weighted whole of the yield of all concealed node and produces a yield by converting the weighted total using its activation function  $g$ . In time series analysis,  $f$  is usually select as logistic function and  $g$  as an identity function. For  $p$  input nodes,  $q$  concealed nodes, one yield node and predispositions at each concealed and yield layer, the entire number of parameters (weights) in a very 3 layer feed-forward neural system is  $q(p + 2) + one$  [9]. For a univariate time series foretelling issue, pervious perceptions of a given variable serve as the input variables. The neural system model endeavor to delineate the subsequent equation

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-p+1}, w) + \varepsilon_{t+1} \quad (2)$$

Where  $y_{(t+1)}$  relates to the perception at time  $t + 1$ ,  $p$  is that the number of slacked perception,  $w$  is that the vector of network weights and  $\varepsilon_{(t+1)}$  is the mistake term at time  $t + 1$ . Hence, the neural system has done like non-linear autoregressive model [6]

### III. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization was discovered by social psychologist James Kennedy and American electrical engineer Russell Eberhart, it is a sort of swarm intelligence-based computational technique that inspired by the social behavior of flocks of birds as they move from one place to another through the searching idea for food. which aims to obtain convergent solutions to the optimal solution depending on search and repetition. It is applied in many fields like neural system, Biomedical applications, clustering, distribution network, ect.

PSO has been generally used to solve problems related to optimization and problems that change over time. The idea of the algorithm is based on a swarm of elements spreading in a limited research area and moving randomly to search for the optimal solution in this region. Generally, as the number of swarms increases and the search area becomes smaller, finding the optimal solution becomes easier and faster, Vice versa, As the number of elements is smaller and the search field becomes larger, finding the optimal solution became more difficult[7].

### IV. USING PSO TO OPTIMIZE ANN

The key drawback to try and do the statistic foretelling using an Artificial neural network is to reconstitute the measured information and establish the performance relationship between preliminary observation data and the next data.

Training process begins by entering data into the network, these data composed of two parts: the first part represents the independent set (Input variables) and the second part represents the dependent set (target variables), these two sets together are modeled by a matrix, which represents the inputs data of the network. The matrix form can be written as follows:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_p & T_{p+1} \\ x_2 & x_3 & \dots & x_{p+1} & T_{p+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n-p} & x_{n-p+1} & \dots & x_{n-1} & T_n \end{bmatrix} \quad (3)$$

It should be noted that all the input coaching information ought to be normalized initial, appropriate embedding dimension ought to be calculated; then coaching and testing samples is established and normalized. The normalized methodology is shown in eqn. (4)

$$\hat{x} = 2 * ((x - x_{\min}) / (x_{\max} - x_{\min})) - 1 \quad (4)$$

Where  $x$  is measured data and  $y$  is normalized data,  $x_{\max}$  and  $x_{\min}$  are the maximum and minimum values of the measured data. In many cases, the prepared data for the training contains high values. So, it is always advisable to normalize the inputs and outputs of the network to be within the range  $[-1, 1]$ , and this normalization has many advantages such as:

- 1 All inputs become convergent values and therefore does not dominate the entrance to another.
2. The normalization values make training faster [6]. The following steps can describe the general steps of training ANN using PSO.

#### A. Create Initial Particles

To optimize the initial weights of the neural system, all layers of feed-forward neural system weights and bias ought to be encoded as a particle model (Fig. 2.). because of the neural network weights and bias parameters are a complexity continuous improvement drawback, it's common to use ancient real encryption technique, and this additionally can have an effect on the accuracy and process potency of evolution rule, as a result of this encryption technique suits for giant span and high exactitude particle swarm improvement. All layers of neural system weights are encoded to PSO Individual as per request Every particle consists of a vector of  $N$  weights were generated randomly from the uniform distribution in the range  $[-1, 1]$ . The total number of weights of  $(p, q, 1)$  ANN model is  $q(p+2)+1$ .

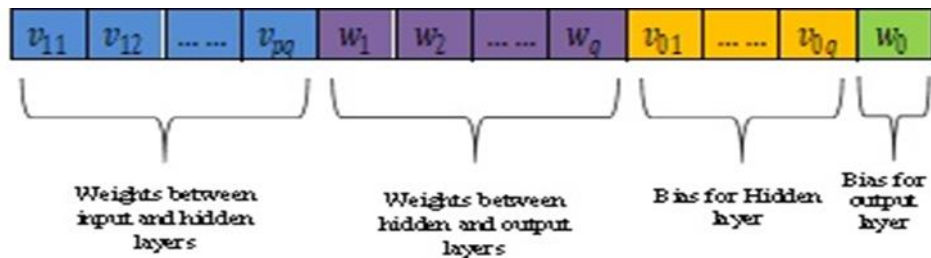


Fig. 2. The particle structure

## B. Fitness Evaluation

The value of fitness function is that the mutual value of the error sum of squares. The fitness function is considered a formula (5).

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (5)$$

The current best fitness is done by particle p is about as pbest. The pbest with best value is about as gbest and this value is keep

## C. Particles Movement

At every iteration, each particle moves towards its own best position and towards the better particle of the swarm found up until this point. The speed changes as indicated by equation (6).

$$V_i^{n+1} = w * V_i^n + c_1 * r_1 (P_i^n - x_i^n) + c_2 * r_2 (P_g - x_i^n) \quad (6)$$

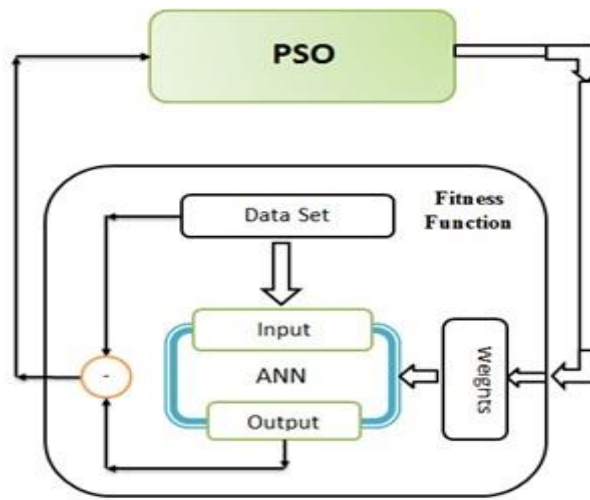
Where n is that the iteration sequence of the particle i; c1 and c2 are certain consistent parameters called speeding up coefficients which are answerable for controlling the greatest step size, r1 and r2 are irregular numbers between (0, 1); w is that the inertia weight;  $V_i^{n+1}$  is particle i's speed at cycle n+1;  $V_i^n$  is particle i's speed at cycle n;  $x_i^n$  is particle i's position at cycle n;  $P_i^n$  is the recorded individual best situation of the swarm (pbest);  $P_g$  is the global best particle (gbest). At long last, the new position of particle I;  $x_i^{n+1}$ , is determined as appeared in (7).

$$x_i^{n+1} = x_i^n + V_i^{n+1} \quad (7)$$

To improve the performance of PSO, To make the performance of PSO more efficient, the parameters are often adjusted. for instance, the constant w are exchanged by following equation (8).

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \quad (8)$$

The inertia term plays a vital role on the performance of PSO because it allows equilibrium between global and local search throughout optimization method. Fig.3. shows how to train ANN using ANN.



**Fig.3.** How to train ANN using PSO

#### **D. Stopping Condition**

The purpose of training artificial neural networks using PSO is to get a balance between the correct responses to training samples as well as good responses for new entry samples (i.e. the balance between remembering and forecasting). This training process is not supposed to continue until we get to the smallest value of MSE, but it can be determined by a stopping condition.

So, the data were randomly split into two groups during the training process (training group and testing group), these two groups are independent and there is no correlation between them.

The algorithm will continue in training and improving the weights, it depends choose weights on the output of the use of the test set error, in each iteration, the error value of the square of the group training and a test account and as long as the error to set the test decreases the training process will continue, and when it begins to increase the network remembering the training samples and after so it will lose its ability to predict, and at this point ends the training process. Fig. 4. shows the best time to stop the training process.

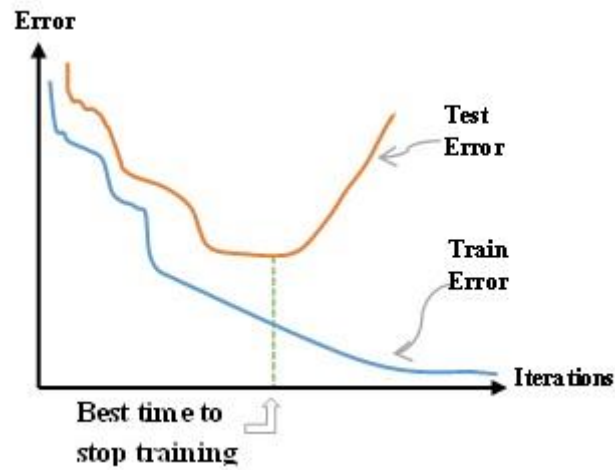


Fig. 4. The best time to stop training

In this paper, we stopped training process at the end of the evolutionary algorithm iterations and then choose the specific weights group that gave less square error for the group to be a test weights

## V. EXPERIMENTS AND DISCUSSION

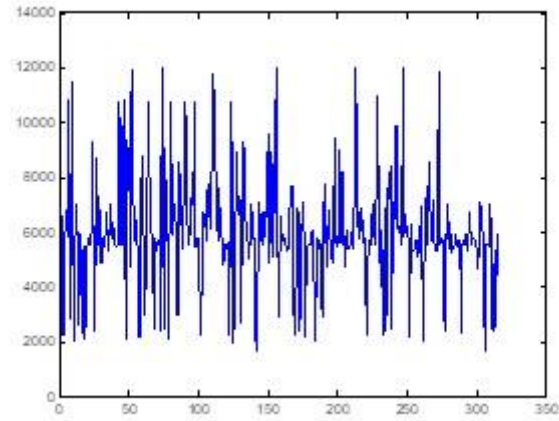
### A. Data

Two data sets the packed cement data and the towels data are utilized in this paper to prove the efficacy of the ANN optimized by PSO to forecast the demand.

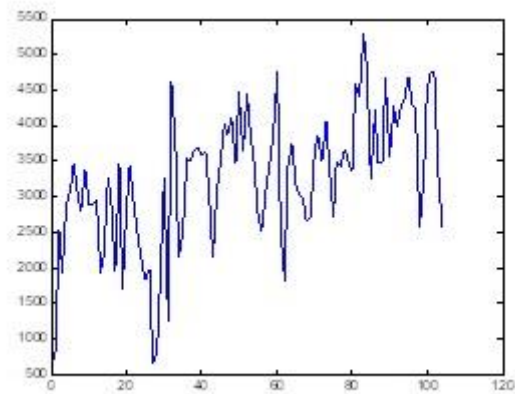
The packed cement series we consider contains the weekly demand data for packed cement from 02.01.2007 to 05.01.2013, giving a total of 315 observations. The data are plotted in Fig. 5.

The towels series contains the weekly demand information for hand towels, that are created by public Companies of prepared clothes. The data are plotted in Fig. 6. The data set has 104 observations, relating to the time of 12.01.2011 to 09.01.2013





**Fig. 5.** Time series plot for packed cement



**Fig.6.** Time series plot for towels

To determine the prognostication performance of various models, every data set is split into two samples of coaching and testing. The coaching data set is utilized only for model development and afterward, the test sample is utilized to assess the built-up model. Table I show data compositions for the two data sets

**Table I.** Sample size Sample compositions in two data sets

Series	Sample size	Training set size	Test set size
Cement	315	240	75
Towels	104	78	26

## B. The network architecture

As previously determined in this paper, a three-layer feed-forward neural network model has been utilized for this paper, also the strategy and personality function have been utilized as the activation function for the concealed nodes and yield node respectively. In designing ANN for forecast, one must determine the following variables:

- number of input nodes  $p$
- number of hidden nodes  $q$

The choice of these parameters is essentially trouble dependent. Although there exist many different methods which include the pruning algorithm, The polynomial algorithm and the genetic set of rules for locating the greatest structure of an ANN, these methods are usually quite complicated in nature and are tough to implement. In this study, in order to optimize the structure of every network genetic algorithms are used [3]. The most number of nodes within the input layer is ten and the most extreme number of the nodes in the concealed layer will be fifteen. to decide the best structure for every series in terms of the number of the input nodes and the number of the nodes within the hidden layer, it was utilized the learning technique of the system and transfer perform, the capacities of the optimizing genetic technique obtainable in Neuro Solutions computer code. The best structure among many runs for packed cement data was [7,9,1] and for towels was [10,5,1].

### C. Training and Forecasting

Particle swarm intelligence could be a population-based mostly search strategy that finds a better answer in the pursuit space. The error surfaces for neural systems are a lot of irregulars and include many irregular geometries. Consequently, for this paper. ( $c1=1.5$ ,  $c2=1.5$ ,  $wmax=1.7$ ,  $wmin=0.6$ ,  $Nparticles=100$ ,  $Ngen=1000$ ) are used for coaching the neural network. As indicated earlier, every particle is related to 2 parameters: position and speed. Initially, we have a tendency to haphazardly distribute particles in many nodes of the search area. every particle is related to a position and a speed. of these nodes within the scope represent a neural network pattern. The optimization experiment was run fifty times with totally different initial values for location and speed to record average and therefore the best value for performance measures.

The prediction capacity has been evaluated according to two popular performance metrics: the mean square error MSE and also the mean absolute deviation MAE. The MSE measures the performance of namely vi six The MSE measures the general performance of namely 6 weeks ahead and 10 weeks ahead respectively. Each table from (II-III) indicates the best MAE and MSE based on fifty runs for various coaching methods concerning a prognostication information series, which are. The MSE and MAE were determined to evaluate the consistency of the performance of various techniques.

**Table II.** Forecasting comparison for Packed Cement data

Method	6 points ahead		10 points ahead	
	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
PSO	$7.992 \times 10^5$	764.64	$1.248 \times 10^6$	965.5
Levenberg-Marquardt	$9.382 \times 10^5$	854.44	$1.566 \times 10^6$	1068.5

**Table III.** Forecasting comparison for towels data

Method	6 points ahead		10 points ahead	
	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
PSO	1868.27	108.31	2801.59	127.80
Levenberg-Marquardt	2169.65	113.19	3060.82	130.33

## D. Discussion

After displaying the results, we show that the ANN-based coaching algorithm gives higher prognostication ability according to the best MAE and MSE across the two prediction horizons. In packed cement forecasting, the best MSE value of  $7.992 \times 10^5$  was obtained in case of PSO as compared to the best MSE value of  $9.382 \times 10^5$  for Levenberg-Marquardt technique for 6 weeks duration. With the PSO trained algorithm, We obtained 17.39% and 25.27% in MSE across the neural network trained for posterior proliferation for 6 weeks and 10 weeks ahead of expectations. In terms of MAE, PSO provided 10.5098% and 9.63 % decrease as compared to usual back propagation algorithm for 6 weeks and 10 weeks respectively.

In towels forecasting, the best MSE value of 1868.27 was obtained in case of PSO comparing to Levenberg-Marquardt algorithm for six weeks period which obtained 2169.65 for best MSE value. With the PSO trained algorithm, we obtained 16.13% and 9.25% increment in terms of MSE over the standard back-propagation trained neural network for 6 weeks and 10 weeks ahead forecasts respectively. In terms of MAE, PSO gave 4.50% and 1.941% decrement as compared to the common back propagation technique for 6 weeks and 10 weeks respectively.

Further, neural system model dependent on PSO uniformly gave preferable foretelling precision than Levenberg-Marquardt concerning mean and the best value of performance metrics over all the two anticipating horizons (As seen in Fig.7). it's been seen that the mean values of MAE and MSE were in every case high for PSO based coaching technique as compared to common back propagation coaching technique. The foretelling capacity of all the coaching techniques disintegrated with the expansion in determining periods which is self-evident. As a result of this reason neural system based models are favored for short term forecasting.

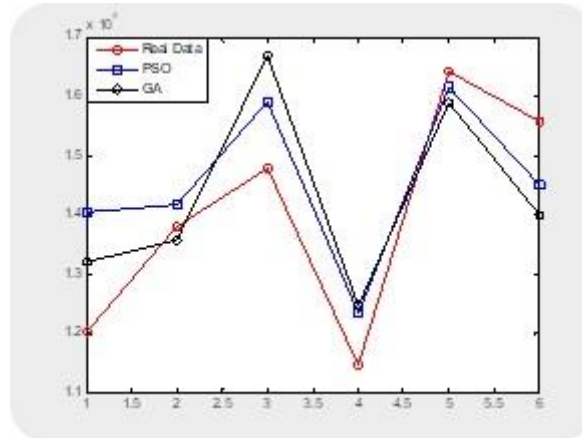


Fig.7. PSO and BP forecasting of packed cement for 6 weeks

## VI. CONCLUSION

Time series analysis and request prediction are dynamic research regions over the past few decades. The demand forecasting is fundamental to many decision processes and hence the research for improving the effectiveness of forecasting models and algorithms has never stopped. In this paper, attempts were created to evaluate PSO's predictive ability to coach a multilayer feed-forward neural system with a genuine demand time-series information. for the train, the neural network, PSO alongside Levenberg-Marquardt technique was used .we show that the PSO-based coaching algorithm provides higher predictability according to the best MAE and MSE through the two prediction horizons and for each data series.

### Acknowledgments

The authors are very grateful to the Northern technical University /Technical College of Management and to the university of Mosul/ college of Computer Science and Mathematics for their provided facilities, which helped to show this work in this format.

### References

- [1] Moghaddama.A , Moghaddamb.M and Esfandyari c.E, "Stock market index prediction using artificial neural network " , Journal of Economics, Finance and Administrative Science 21 (2016) 89–93.
- [2] D . Fogel, Evolutionary Computation: The Fossil Record. Piscataway NJ: IEEE Press, 1998.
- [3] D. Goldberg, Genetic Algorithms in Search Optimization & Machine Learning. Addison-Wesley. Reading, 1989.
- [4] D. Srinivasan, and T.H.Seow, "Evolutionary Computation", In: CEC 2003.December 8-12. vol. 4. pp. 2292–2297. Canberra. Australia, 2003.
- [5] Tealab.A , "Time series forecasting using artificial neural networks methodologies: A systematic review " , Future Computing and Informatics Journal ,Volume 3, Issue 2, December 2018, Pages 334-340.

- [6] Datilo.P , Ismail.Z and Dare.J , “A Review of Epidemic Forecasting Using Artificial Neural Networks” , International Journal of Epidemiologic Research , 2019 Summer;6(3):132-134
- [7] Dian Palupi Rini , Siti Mariyam Shamsuddin and Siti Sophiyati Yuhaniz, "Particle Swarm Optimization : Technique , System and Challenges " , University Teknologi Malaysia , International Journal of Computer Applications ( 0975 – 8887 ) Volume 14– No.1 , January, 2011
- [8] Wang.D, Tan.D• and Liu.L , “Particle swarm optimization algorithm: an overview “ , Published online: 17 January 2017 © Springer-Verlag Berlin Heidelberg 2017 , <https://doi.org/10.1007/s00500-016-2474-6>
- [9] K. Girish, T. Parimala, and K. Ruppa, "PSO Based Neural Network for Time Series Forecasting", Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, 2009, pp. 1422–1427.
- [10] Wiles.P and Enkeb.D , “Nonlinear Modeling Using Neural Networks for Trading the Soybean Complex “ , Procedia Computer Science , Volume 36, 2014, Pages 234-239 <https://doi.org/10.1016/j.procs.2014.09.085>
- [11] Nasser.I, Al-Shawwa.M and Abu-Naser.S , “A Proposed Artificial Neural Network for Predicting Movies Rates Category “ , International Journal of Academic Engineering Research (IJAER) ,ISSN: 2000-001X , Vol. 3 Issue 2, February – 2019, Pages: 21-25
- [12] Kukreja.H , et al. , “ An Introduction to Artificial Neural Network” , International Journal Of Advance Research And Innovative Ideas In Education , Vol-1 Issue-5 2016 ,ISSN (O)-2395-4396
- [13] Gupta.A, Agarwal.A and Garg .R , “ Analysis of Back Propagation Neural Network Method for Heart Disease Recognition” , International Journal of Organizational and Collective Intelligence (IJOICI) 9(2), DOI: 10.4018/IJOICI.2019040104 , 2019