A Prediction of Grain Yield Based on Hybrid Intelligent Algorithm

Ibrahim Ahmed Saleh¹, Wasan Abdallah Alawsi², Omar Ibrahim Alsaif³, Khalil Alsaif⁴

¹ University Of Mosul, Software Engineering Dept., Iraq, i.hadedi@uomosul.edu.iq
² University Of Al-Qadisiyah, College Science, Iraq, wasan.alawsi@qu.edu.iq
³ Northern Technical University, Mosul Technical Institute, Iraq, omar.alsaif@ntu.edu.iq
⁴ University of Mosul, Computer Science Dept., Iraq, khalil_alsaif@hotmail.com

Abstract. The prediction is most important goals in economic quantitative studies, it basis in design and plan future economic policies properly process over forecasting accuracy. This paper is aiming at the problem salp swarm algorithm (SSA) for predicting grain yield is prone to fall into the local optimal problem. An improved SSA is proposed with combine with back propagation neural network. Using the different advantages of SSA algorithm in global search capabilities, combining the two for further optimize the weight, improve the accuracy and robustness of the grain yield prediction model. The specific implementation is selected from 1963 to 2013.

These methods are used to define agricultural datasets that supports crop growth decision for grain product and its influencing factors were tested as a data set. The results show that, the improved salp swarm optimization can be classified as a good predict tool for the domestic food production trend in recent years compared with the SSA. This paper briefly introduces three artificial methods BP neural networks, SSA and improved SSA optimization algorithm. The natural behavior of salp, barrel-shaped plankton that are mostly water by weight optimization and combined with mixed-group of intelligent algorithm are simulated. The simulation results of grain production prediction illustrate that the predict precision of the improved SSA is much higher than of both conventional BPNN and SSA techniques and it’s very efficient and practicable.

Keywords: grain yield; prediction; Salp swarm optimization; neural network, back propagation algorithm

1. Introduction

Grain is given the high position in the Iraqi food basket, which in turn is formed 35% of the per capita income for Iraq yield, it is one of the basic guarantees for human survival. Food is related to the country’s economic prosperity and social harmony [1]. In order to ensure the sustainable development of the agricultural market, determine the complex relationship between food production and its influencing factors. The effectively grasp and regulate the influential factors that can be intervened, and establish a feasible, accurate and easy-to-realize food production dynamics Prediction models are imperative [2]. Due to the characteristics of randomness, non-linearity and dynamic nature of the food system, it is difficult for traditional prediction methods to effectively grasp and predict it. As one of the most successful optimization method currently applied, the outstanding advantage of these methods they have optimal solution for nonlinear mapping and are often used to solve complex nonlinear problems [3]. SSA was proposed to solve different kinds of optimization issues and provide various solutions to different problems [4].

Literature [5], summarized such improvements in feed-forward neural networks in more than 20 years. As a new type of biological evolution algorithm, particle swarm algorithm relies on the information exchange mechanism between its population particles and has good global optimization ability. This feature can be used for the shortcomings of BP neural network. However, particle swarm algorithm is easy to apply for the optimization process that specify local convergence and precocity [6].

Literature [7], summarized for improved back propagation neural network with particle swarm optimization (PSO) by introducing mutation operation and adaptive adjust of inertia weight, the problem of easy to fall into local optimum, premature, low precision and low later iteration efficiency of PSO are solved. The results of grain production prediction show that the predict accuracy the method is effective and feasible.
While literature [8], used Artificial Bee Colony with weighted based Fuzzy Clustering algorithm to predict a district for them country. The researchers are cropped prediction analysis consider as high yield area and used many factors. They improved hybrid artificial bee colony with weighted based fuzzy clustering algorithm yields better performance than other clustering algorithm like k-means and k-medoids with high accuracy.

In this paper three methods for solve this problem are applied, the back propagation neural network (BPNN), salp swarm algorithm and combining between two algorithms. The proposed algorithm called improved salp swarm algorithm or hybrid swarm intelligent algorithm. This improved algorithm could describe as follows [9].

When the salp swarm algorithm falls into convergence, it can basically cover all extreme points during the establishment of the BP neural network prediction model, making its fitting ability stronger.

2. Salp Swarm Algorithm (SSA)
The SSA is a new nature-inspired optimizer using a method to simulate the swarm behavior of salps in nature. SSA algorithm detects, satisfies verify and strengthening tendencies that made it interesting for developing ELM training tasks. The SSA can be looked as a flexible, capable, simple, and understudied easy [8][9]. Local convergence optimal is very critical, so the salps algorithm are updated their position vectors gradually considering dynamic crowd of agents for other salps. The salp swarm optimization has iterative natural, that lead to save managing other members of swarm for better areas. The salps vector contain leader and followers which should be update their location vectors. The leader of swarm will determine the direction of food source (F) while the movement of all followers will be moved towards leader directly or indirectly [8] [10]. Figure(1) explains the salp chain movement.

![Salp Swarm Algorithm](image)

**Figure 1.** salp’s chain movement and the model of leader and follower.

The salps population (X) contain N agents with d-dimensions, hence the matrix (N*d) can illustrated in equation (1):

\[
X_i = \begin{bmatrix}
X_{(1,1)} & X_{(1,2)} & \cdots & X_{(1,d)} \\
X_{(2,1)} & X_{(2,2)} & \cdots & X_{(2,d)} \\
. & . & \cdots & . \\
. & . & \cdots & . \\
X_{(1,N)} & X_{(2,N)} & \cdots & X_{(N,d)}
\end{bmatrix}
\]

(1)

For this algorithm, the leader and followers are updated there location based on equation (2), [7]:
\[ x^j_I = \begin{cases} 
F_j + c_1 \left( (u_{bj} - l_{bj})c_2 + l_{bj} \right) & c_3 \geq 0.5 \\
F_j - c_1 \left( (u_{bj} - l_{bj})c_2 + l_{bj} \right) & c_3 \geq 0.5 
\end{cases} \tag{2} \]

Where:
- \( x_{1j} \) denotes the leaders’ position.
- \( F_j \) represent to best solution of food source in the \( j \)th dimension
- \( u_{bj}, l_{bj} \) are the upper and lower bounds in the \( j \)th dimension.
- (c1, c2 and c3) are random values.
- c1 is the main parameter of algorithm, which have significant role for the performance of SSA, It is consider as the only parameter that that calibrate and manages the balance between exploration and exploitation processes. This parameter depends on the iteration number that permits high exploration proportions as in equation (3) [9][10]:

\[ c_1 = 2e^{-\left( \frac{4t}{T_{max}} \right)^2} \tag{3} \]

Where:
- \( t \) iteration, while (\( T_{max} \)) maximum number of iterations. By increasing iteration count, this parameter decreases. \( c_2 \) and \( c_3 \) generated in the period \([0, 1]\). As a result, it can manage to put more approve on the variegation tendency on primary steps and put more confirm on intensity movement in last periods of optimization. The followers positions can be corrected using equation (4) [11]:

\[ x^I_j = \frac{x^I_j + x^{I-1}_j}{2} \tag{4} \]

Where \( i \geq 2 \) and \( x_{ij} \) is the location of the \( i \)th follower salp at \( j \)th dimension. The pseudo-code of SSA is expressed in Algorithm 1 as shown in Figure (2).

**Figure 2.** Pseudo code of Salp Swarm Algorithm
3. Proposed Approach

In this paper, a crop dataset was collected from which included the Ministry of Agriculture – Department Planning and follow-up, Agricultural Statistics and Manpower Division, and - Supply Card Department. The dataset is preprocessing by redundancy by particle data filtering, and three algorithms were applied using Matlab software to predict which area will produce high yields for specific crop in particular season.

3.1 Back Propagation Neural Network Algorithm

(BPNN) is a multi-layer feed forward neural network, proposed since 1986, So far its application has spread to various fields for arbitrary nonlinear programming, BPNN can effectively fit and predict. It’s mainly divided into two behaviors: forward transmission of the input signal and error reverse transmission of signals [10].

During this period, the network model tends to be optimized. The main steps in establishing a BP neural network are weights and thresholds. Optimization is occurred through training on a large number of data sets establishes and generates a network model, as shown in figure (3).

![Figure 3. Back propagation topology network](image-url)

Here, $X_1, X_2, \ldots, X_n$ denoted neural network input amounts of, $Y_1, Y_2, \ldots, Y_n$ are the predictive values, $w_{ij}$ and $w_{jk}$ weights between input- hidden layer and hidden layer-output layer respectively the training process as shown in figure (4)[11].
Intermediate unit is usually set to a single hidden layer, but when the input layer have many elements, the multiple set hidden layers will be a good choice for fitting nonlinear functions better performance, more accurate predictions, and longer training time. Therefore, for more complex mapping relationships can use multiple implicit ways to build a network. In this experiment, the network structure is more accurate in predicting food production, but it is also corresponding increased runtime.

3.2 SSA for Grain prediction

There are some essential steps should be followed to predict by SSA behavior:

Initial parameter for SSA number of swarm:

\( (N)=200, \text{dim}=5, \text{lb}=0, \text{ub}=1, c2=c3=0.5, \text{Tmax}=500; \)

Load the data set generation values of the algorithm.

The fitness computation process which in Euclidian distance is carried out for each site visited by a (SSA) using clustering algorithm implemented for the crop dataset to predict the district as shown below.

\[
D_{ij}^2 = \sum_{p=1}^{n} (X_{pi} - X_{vj})^2 \quad (5)
\]

Amid distance of the salps based on upper and lower bound of variables.

Locations update step for the exploration and centroid.
3.3. Technique of Hybrid Optimization Algorithm

The main problem of swarm algorithm lies in local or global convergence. In this paper (SSA) used to improve the inertia weight mainly optimize from two aspects. Using particles always maintain a good search ability which has fallen into convergence for points. A new method of assigning positions to make them free from the constraints of extreme points.

In the specific optimization process of SSA algorithm the weight (w) is improved with obvious effect and need to adjust quickly and effectively [12]. A larger weight can ensure that the SSA algorithm has strong search capabilities, it can make the entire algorithm quick converge, but it is easy to find optimum values [11] [12].

Combining the current research on the inertia weight of the SSA proposed a strategy to change the inertia weight following the position vector that illustrated in equation (6) [9].

\[ w = w_{\text{max}} - \frac{|x_0 - g_i|}{x_{\text{max}} - x_{\text{min}}} \times (w_{\text{max}} - w_{\text{min}}) \]  

(6)

Where: wmax and wmin are the maximum and maximum values of the inertia weight respectively. Small values of xmax and xmin are the boundary values of particles in space, while x0 and gi the initial position of the ith particle and the optimal position of the current group. During the movement of swarm, the salps are gradually approaching the target value, then (x0- Gi) gradually becomes larger. According to equation (5), (w) shows a trend of linear decrease. Simulation tests prove that this method relatively speeds up the algorithm's convergence speed and prediction accuracy, which lead to improving the efficiency of the SSA algorithm. Predictive model of BP neural network optimized by SSA can be described as the following steps [14] [15]:

1) Initialize the parameters of the BP neural network.
2) Initialize SSA optimization algorithm to determine the population number and initial value.
3) Load the data, then (BP) neural network makes iteration over the data set generation, then returned error function is used as the adaptive value of the optimization algorithm.
4) The optimization algorithm updates the particle position according to the fitness value illustrated in equation (5) and will update after the value is returned to the BP neural network as the weight of the next iteration and threshold.
5) Substituting the weights and thresholds obtained in step (4) into step (3), Repeat iteration to the end condition of the optimization algorithm.
6) Return the final particle position to the BP neural network as the initial value of prediction model.
7) BP iterates the optimal network structure.
8) Substitute test data and output prediction results.

4. Experiment Results and Discussion

Training models in food production prediction require a lot of training data. In order to obtain better prediction results. The grain production data from 1963 to 2013 and eight factors that affect production used as data sets. The data comes from the website Iraqi ministry of agriculture (planning department), department of agricultural statistics and the central Agency for statistics and information technology, after comprehensive consideration for many factors affect to grain yield include [13] [14]:

- Sown area with grain.
- Total irrigated agricultural area.
- Manpower.
- Total amount of agricultural fertilizers.
- Total power of agricultural machinery owned at end of the year.
- Total areas guaranteed rain.
- Total rainfall areas and amount of rain.
- The total agricultural area is not guaranteed rain.
A total of eight indicators were used as the input values for the affected area, after causing incomplete function mapping, a total of \((50 \times 8)\) data input matrices and \((141 \times 1)\) data output matrices are formed. In addition, due to the unequal data acquisition methods, the results predicted and analyzed by the network model generated by the above data may not be consistent with the base and economic information. In the next few years, larger data set needs to be established for fitting prediction. For all data samples, consider the impact of noise, abnormal data, or inconsistent information acquisition methods. Before the neural network training, perform dimensionality normalization processing, so that all sample data is converted into \([0, 1]\) Standard data [14]. An improved maximum and minimum method is used to process it, and the calculation formula is as follow:

\[
    x_i = 0.1 + 0.9 \times \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad i \in [1, n]
\]  

(7)

The results prove that the normalized data has a stronger prediction effect, and avoids the phenomenon that a relatively consistent data dimension exists in the sample and returns to zero.

4.1 Prediction results and comparison processes

4.1.1 Prediction of BP neural network

When BP neural network is used to establish a prediction model after number of training when the number of network nodes is taken, the BP neural network has the best fitting ability to establish a prediction model. The samples with data numbers \((1\ to\ 121)\) are used as the training set, and the remaining 20 samples are used as the test set. The first \((8)\) items of the sample will be input to the network, and the last one is output from the BP neural network. The prediction result and actual grain of BP network is shown in figure (4).

![Figure 5. Prediction results of BP neural network](image)

It can be seen from figure (5), that the predicted output results generally conform to the overall trend of the expected output, but the errors between the predicted value and the expected value are large at some points. The reasons are as follows:
1) The BP neural network is trapped in local convergence during the training process, and the best weights and thresholds are not found, so the result is a model that is not accurate enough.

2) In the test set, the mutual influence relationship between all input data is more complicated, and there are many factors influencing the grain output itself. There are large differences in the effects of natural factors and scientific and technological factors on the grain output, and the relationship between such data was not handled well when the data set was established.

3) The use of neural networks to make predictions requires a large amount of data as test samples. Under the current conditions, a large number of complex samples have not been formed, resulting in large fluctuations in some data during prediction. For example, over the years, the food disaster situation and the arable land disaster area are prone to under fitting.

4.1.2 Prediction of SSA algorithm
Salp Swarm Optimization is based on K-means clustering to analyze the district wise crop prediction, it can be seen from figure (6) the results of this algorithm.

![Figure 6. Prediction results of SSA algorithm](image)

The predicted output optimized by the SSA algorithm is significantly better than the result of figure (3), but the predicted and expected values at some points are still obvious error value, and the fluctuation of the overall prediction result is large. This is because once SSA finds a better optimal value in the optimization process; other particles will quickly converge to this point.

4.1.3. Prediction of improved BP neural network
It can be seen from figure (7), that the prediction effect of improved BP neural network optimized by SSA is significantly better than the previous two prediction models, and the prediction is more accurate. The overall prediction result is consistent with the line chart trend of the expected result. The prediction effect of large fluctuation data is obviously better than the basic BP and SSA models.
Figure 7. Prediction results of improve SSA

The prediction results of the above three methods are compared. The prediction error (the absolute value of the difference between the predicted value and the expected value) is used as the standard for the accuracy. The comparison results are shown in table (1).

| Network structure | maximum error | Average error (%) |
|-------------------|---------------|------------------|
| Bp                | 18.9          | 8.48 ± 0.17      |
| SSA               | 14.7          | 4.53 ± 0.15      |
| Improve-SSA       | 5.88          | 2.42 ± 0.096     |

From Table 1, it can be confirmed that the improved SSA is obviously more accurate in the prediction of food production. At the same time, the optimization ability of SSA is significantly stronger than the basic Bp. The BP neural network optimized by SSA give non-linear model which has better fitting ability, less volatility of prediction results, stronger ability to resist local convergence, and achieve the expected effect.

5. Conclusions

At all agricultural statistical issues it is very difficult to accurately predict grain yield production, in this paper three algorithms are suggested.

1) The inertia weight update method in the SSA algorithm has been improved. The improved hybrid algorithm of SSA has strong global search performance and the ability to resist local convergence. It has played a very important role in the optimization of BP neural network weights and thresholds.

2) A feasible grain production forecasting model is established by using the existing grain production data. Through this model, the grain production can be predicted and analyzed for the next few years.

3) There are many factors that affect the grain production. The forecasting model can be used to perform some analysis one by one to determine the size of each factor's influencing factor. So the training data is improved to enhance the accuracy of the forecast.

4) The improved BP algorithm has a higher time complexity when training the model. In the future work, the search operator can be improved to enhance the search efficiency of the algorithm.
6. References

[1] Shaibu A. S.*, Adnan A. A. and Umar I. R. "Predicting grain yield of maize using drought tolerance traits", African Journal of Agricultural Research Vol. 10(33), pp. 3332-3337, 13 August, 2015 , DOI: 10.5897/AJAR2015.9561 . 2015

[2] Ibrahim RA, Oliva D, Ewees AA, Lu S. "Feature selection based on improved runner-root algorithm using chaotic singer map and opposition-based learning". International conference on neural information processing, Springer, Berlin, pp 156–166. 2017

[3] Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM, Salp swarm algorithm: "a bio-inspired optimizer for engineering design problems". Adv Eng Softw 114:163–191. 2017.

[4] El-Fergany AA. "Extracting optimal parameters of poem fuel cells using salp swarm optimizer". Renew Energy 119:641–648. (2018).

[5] Yok Hack, Abraham A, SON, ELV. "Meta-heuristic disaffects feed for world neural networks": over view decades for research [J]. Engineering Application Artifact Intelligence, 60: 97-116. 2017.

[6] Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems". Advances in Engineering Software. (2017), Volume 114, December 2017, Pages 163-191, https://doi.org/10.1016/j.advengsoft.2017.07.002.

[7] Zhang Liguo, Liu Jiangtao, and Zhi Lifu," Research on Grain Yield Prediction Method Based on Improved PSO-BP” TELKOMNIKA Indonesian Journal of Electrical Engineering Vol. 12, No. 10, October 2014, pp. 7404 ~ 7411 DOI: 10.11591/telkomnika. v12i8.5369.

[8] Surya P. and Aroquiaraj I.Laurence, “Crop Prediction Analysis in North western Zone of Tamilnadu using Artificial Bee Colony with Weighted based Fuzzy Clustering” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019

[9] Abusnaina Ahmed A, Ahmad Sobhi, Jarrar Radi. And Majdi Mafarja "Training Neural Networks Using Salp Swarm Algorithm for Pattern Classification", ICFNDS 2018, June 2018, https://doi.org/10.4751/123_4.

[10] Ibrahim A. Saleh, Omar. I. Alsaiif, SA Muhamed, El Essa " Task Scheduling for cloud computing Based on Firefly Algorithm" Journal of Physics: Conference Series 1294(4), 042004, 2019.

[11] Li Heli, Wang Miao, Li Bo. "Integrated particle swarm optimization for single-objective optimization Law". [J]. Journal of Chongqing University of Posts and Telecommunications. 2018.

[12] Wu Yuming, Li Jianxia. "Non-linear artificial neural network model and its application in corn production prediction". Journal of Henan normal university (Natural Science). 30: 35-38. 2002.

[13] EBERHARTR C, SHI. Y, 2000, Comparing inertia weights and constriction factors in particle swarm optimization, Proceeding of the 2000 Congress on Evolutionary Computation . Piscataway: IEEE computer Society 2000:84-88.

[14] YGOLOVN, R? NNB, CKL. "Big data normalization for massive parallel proceeding data basses". J. Computer Standards and Interfaces, . 54: 86-93. 2017.

[15] Eman A. Abdullah, Ibrahim A. Saleh, Omar I. Al Saif. "Performance Evaluation of Parallel Particle Swarm Optimization for Multicore Environment". International Conference on Advanced Science and Engineering (ICOASE) – Duhok, Kurdistan Region – Iraq. 2018