Prediction of soil moisture based on BP neural network optimized search algorithm

Xiaoyu An and Fuxing Zhao*
School of Zhengzhou University of Light Industry, Zhengzhou, China

*Corresponding author e-mail: fuxingzhao@zzuli.edu.cn

Abstract. The accuracy of traditional soil moisture prediction method is low and the training period is long. In this paper, the BP neural network prediction model is studied, and a longicorn beetle search algorithm (BAS) optimized BP neural network prediction method is proposed. Spelman rank correlation coefficient method is used to analyze the correlation between the change of soil moisture and each variable. In this paper, evaporation, ground temperature, precipitation, air pressure, sunshine hours, air temperature and wind speed are taken as independent variables to analyze the Spelman correlation with soil moisture, and the correlation between each variable and soil moisture change is obtained. The longicorn beetle search algorithm is used to optimize the initial weight and threshold of BP neural network, and the prediction model of BAS-BP neural network is established. The soil moisture prediction model of BAS-BP is compared with different prediction models of GA-BP and BP. The results show that the average absolute error and average relative error of BAS-BP are 9.1936 and 0.1333 respectively, which is lower than that of GA-BP and BP model. The shortcomings of long training time and slow convergence speed are overcome by BAS-BP neural network, and the accuracy of prediction is improved.

Keywords: Moisture prediction, spearman rank correlation coefficient, longhorn beetle algorithm, BP neural network.

1. Introduction
Our country has 2,800 cubic kilometers of water resources, but the per capita water resources are less than 1,000 cubic meters, which is a severe water shortage. The annual agricultural water consumption accounts for 65.45% of the total water consumption, of which irrigation accounts for 90% of the agricultural water consumption [1]. With limited resources, constantly improving irrigation technology and mastering a reasonable irrigation time have played a very important role in the growth of farming and the protection of water resources [2, 3]. Therefore, the research on the prediction of soil moisture is regarded as one of the most important topics by researchers [4].

The research on soil moisture prediction was that Li used BP neural network to build soil moisture prediction model in reference [5]. The model had good prediction accuracy. Using the default factor to build the soil moisture prediction model of BP-ANN was proposed by Huang [6], which had high fitness and manoeuvrability. The artificial neural network was used to build the mathematical prediction model of soil moisture, and the calculated value of the model is in good agreement with the
measured value [7]. Genetic algorithm (GA) was used to optimize BP neural network, and the soil moisture prediction model based on GA-BP algorithm was established, which had the advantages of small absolute error and fast convergence [8]. The soil water content was regarded as a whole system, and the prediction model established by the water content balance principle of the whole soil system under the condition of constantly changing input and output. This model requires few parameters and was practical [9].

In this paper, the data analysis method of Spelman rank correlation coefficient was used to study the correlation between the change of soil moisture and each variable, and the correlation between soil moisture and each variable was obtained. Using the combination of BAS algorithm and BP neural network, the prediction model of BAS-BP was established to predict soil moisture. The prediction results were compared with the BP prediction model to further verify the effectiveness and fitness of the prediction model.

2. Research objects and experimental tools

2.1. Overview of the study area
This paper is based on the analysis and experiment of the farmland environment in Pingdingshan City, Henan Province, which is located in East China and belongs to the temperate continental monsoon climate, with an average altitude of 400 meters. The farmland environmental variables collected from this area are evaporation, ground temperature, precipitation, air pressure, soil relative humidity (soil moisture), sunshine hours, air temperature and wind speed. The measured data are from March 2019 to September 2019, with 57 days (26.64%) of precipitation from March to September and 157 days (73.36%) of no precipitation. The total data (N) is 214 groups.

2.2. Experimental tools
The database used in this experiment uses Excel to enter various data, and then uses SPSS17.0 statistical software to conduct experimental analysis on the entered data and draw the final experimental conclusion.

3. Analysis method

3.1. Spearman rank correlation coefficient
Spelman correlation coefficient has been widely used in statistics, which is mainly used to eliminate singular values, evaluate the correlation between data, and evaluate the monotonous relationship between variables [10].

The Spearman correlation coefficient is based on the rank of the original data, let (X, Y) be a random variable, and (R, Q) be the rank of the variable (X, Y).

\[
\rho_{X,Y} = \frac{\sum_{i=1}^{n}(r_i-ER)(q_i-EQ)}{\sqrt{\sum_{i=1}^{n}(r_i-ER)^2 \sum_{i=1}^{n}(q_i-EQ)^2}}
\]  

(1)

Where \(r_i\) is the rank of \(x_i\) in the sequence \(\{x_1, \ldots, x_n\}\), and \(q_i\) is the rank of \(y_i\) in the sequence \(\{y_1, \ldots, y_n\}\).

When Spelman’s correlation coefficient is described as the Pearson correlation coefficient between rank variables, for a sample size of \(n\), an original data will be transformed into rank data, and the correlation coefficient \(\rho\) is:

\[
\rho = \frac{\sum_i(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i(x_i-\bar{x})^2 \sum_i(y_i-\bar{y})^2}}
\]  

(2)
In this paper, Spearman correlation analysis is used to calculate the correlation between soil moisture and environmental variables. According to the Spearman correlation analysis, eliminate the data that are not related to the change of soil moisture, or the data with less correlation. The results after elimination are used as the input variables for the prediction of soil moisture.

3.2. BP neural network

The BP neural network inputs the signal through the input layer, and then carries on the algorithm processing in the hidden layer. If the processed data is compared with the real data, if the difference does not meet the set precision requirements, it will enter the back propagation [11, 12]. The weights and thresholds of neurons in each layer will be changed, and the weights and thresholds will be changed continuously through the above process. Until the minimum difference or the preset number of training is met [13]. The process is as follows:

1) Parameter initialization. M, Q, L are set to the number of nodes in the input layer, hidden layer and output layer of the network, and then initialize the weights and thresholds between each neuron layer.

2) Forward propagation algorithm. The hidden layer output is shown in formula (3).

$$H_i = f \left( \sum_{j=1}^{M} \omega_{ij} - a_i \right), i = 1,2,\cdots,q$$

Where:
- $H_i$ represents hidden layer output;
- $j, i$ represent input layer and hidden layer, respectively;
- $\omega_{ij}$ represents the connection weight between the input layer and the hidden layer;
- $a_i$ represents the hidden layer threshold;
- $f$ denotes the hidden layer excitation function.

The output of the output layer is shown in formula (4).

$$O_k = \sum_{i=1}^{q} H_i \omega_{k i} - b_k, k = 1,2,\cdots,L$$

Where:
- $O_k$ is represented as output layer output;
- $b$ represents the output layer threshold;
- $\omega_{k i}$ represents the connection weight between the hidden layer and the output layer.

3) The error formula (5) between output $O_k$ and expected $y'_k$.

$$e_k = y'_k - O_k, k = 1,2,\cdots,L$$

Where: $e_k$ represents the error between the output of the k-th node and the expectation.

4) The weights and thresholds are updated as shown in equations (6) to (9).

$$\omega_{ij} = \omega_{ij} + \eta H_i (1-H_i) e_k \sum_{k=1}^{L} \omega_{k i} e_k, j = 1,2,\cdots,M$$

$$\omega_{k i} = \omega_{k i} + \eta H_i e_k, i = 1,2,\cdots,q; k = 1,2,\cdots,L$$
\[ a_i = a_i + \eta H_i (1 - H_i) \sum_{k=1}^{L} \omega_{ik} e_j, j = 1, 2, \ldots, M \]  
(8)

\[ b_k = b_k + \eta e_k, k = 1, 2, \ldots L \]  
(9)

Where: \( \eta \) represents learning efficiency.

5) Return to step (4), recalculate the output value of each neuron after updating the weight and threshold. When the output error is less than the set error, keep the current weight and threshold, otherwise repeat the above process until the output error is satisfied.

3.3. Establishment of BAS-BP model

The longicorn beetle search algorithm is proposed by Jiang and Li in 2017 [14]. The algorithm is a heuristic algorithm based on the biological behavior of longicorn beetles looking for food. The algorithm is simple and flexible, avoids the local optimal solution, and is suitable for high-dimensional search [15]. In this paper, the common three-layer structure is adopted, and the soil moisture is used as the output layer. Evaporation, ground temperature, precipitation, air pressure, sunshine hours, air temperature and wind speed are used as input layers. The number of neurons in the hidden layer can be calculated according to the empirical formula \( h = \sqrt{m+n+(1-10)} \) so as to establish a neural network model.

The specific steps of the longhorn beard algorithm are:

① The direction of the head of the beetle is random, so a random k-dimensional unit vector is used to represent the vector from the left beard to the right beard of the beetle, which is normalized.

\[ b = \frac{\text{rand}(k,1)}{\| \text{rand}(k,1) \|} \]  
(10)

Where: \text{rand} () is a random function; k represents the spatial dimension. If the BP neural network model is \( M - N - 1 \), the search space dimension \( k = M * N + N *1 + N + 1 \).

② Create the spatial coordinates of the left and right whiskers of longhorn beetle

\[
\begin{cases}
    x_r = x' + d'b \\
    x_l = x' + d'b
\end{cases}
\]  
(11)

Where: \( x_l \) is the position coordinates of the \( t \)-th iteration of the beetle; \( x_r \) is the position coordinates of the \( t \)-th iteration of the beetle’s left beetle; \( x' \) is the centroid coordinate of the beetle in the \( t \)-th iteration; \( d' \) is the distance between the left and right whiskers.

③ Determine the odor intensity of left and right whiskers

The odor intensity of the left and right whiskers can be determined by the fitness function. The direction of the longicorn beetle is determined by the intensity of the beetle, and the function is the fitness function.

\[ x^{t+1} = x' - \delta * b \sin g[ f(x,t) - f(x_l) ] \]  
(12)
Where: \( \text{sign}(\cdot) \) is the symbolic function. \( \delta' \) represents the step factor at the \( t \)-th iteration;

4 Determine the fitness function
The root mean square error of the measurement is selected as the fitness evaluation function,

\[
\text{fitness} = \frac{1}{N} \sum_{i=1}^{N} (t_{\text{sim}(i)} - y_i)^2
\]  

(13)

Where: \( N \) is the number of samples in the training set; \( t_{\text{sim}(i)} \) is the model output value of the \( i \)-th sample; \( y_i \) is the actual value of the \( i \)-th sample.

5 Calculate the fitness value of the long beetle after moving, and set the step factor:
The search ability of longhorn beetle in a certain area is controlled by the step-length factor, and the initial step-length as large as possible is selected to satisfy the current search area without falling into a local minimum. Use linearly decreasing weight method to make the search more detailed:

\[
\begin{align*}
\delta' &= \text{eta}_d \cdot \delta^{t-1} \\
\delta^{t} &= \text{eta}_d \cdot \delta^{t-1}
\end{align*}
\]  

(14)

Where: \( \delta' \) is the distance between the two whiskers in the \( t \)-th iteration, \( \text{eta}_d \cdot \delta \) is the attenuation coefficient of the distance between the two whiskers, \( \text{eta}_d \cdot \delta \) is the attenuation coefficient of the step length of the two whiskers. At this stage, the step factor setting of BASES has not yet formed a complete guiding theory and method. The initial step length \( \delta = 3 \) and the number of iterations \( n = 100 \) have been determined through multiple experiments.

6 To determine whether it meets the conditions for the end of the iteration and meets the end of the iteration, the output is the optimal weight and threshold. If it does not meet the conditions, repeat 2–6 until the conditions are met.

The neural network learning process is:

7 Network initialization. Randomly select the optimized weights and thresholds of digital connections from the interval \((-1,1)\). First enter \( X \) and expectations into the computer, and then set the maximum number of training times and error accuracy \( \varepsilon \).

8 Calculate the output of each layer

\[
\begin{align*}
h_j(k) &= f\left\{ \sum_{i=1}^{m} [w_{ji} \cdot x_i(k) - \theta_j] \right\}, j = 1, 2, \ldots, h \\
a_o(k) &= f\left\{ \sum_{j=1}^{m} [w_{jo} \cdot h_j(k) - \gamma_o] \right\}, o = 1, 2, \ldots, h
\end{align*}
\]  

(15)

9 Calculation error

\[
e = \frac{1}{2} \sum_{k=1}^{i} [a_o(k) - t_o(k)]^2
\]  

(16)
\[ \delta_o(k) = a_o(k)[1 - a_o(k)][a_o(k) - t_o(k)] \]
\[ \delta_j(k) = a_o(k)[1 - a_o(k)]\delta_o(k)v_{jo} \]

(17)

Where: \( \delta_o(k) \) represents the hidden layer node error, and \( \delta_j(k) \) represents the output layer node error. \( e \) is the error function.

(18) Revised weight

\[ v_{jo}^{N+1} = v_{jo}^N + \eta \delta_o(k)a_o(k) \]
\[ \Delta v_{jo}(k) = -\mu \frac{\partial e}{\partial v_{jo}} = \mu \delta_o(k)a_o(k) \]

\[ w_{ij}^{N+1} = w_{ij}^N + \eta \delta_j(k)x_i(k) \]
\[ \Delta w_{ij}(k) = -\mu \frac{\partial e}{\partial w_{ij}} = \delta_j(k)x_i(k) \]

(19)

Where: \( N \) is the number of iterations, \( \eta \) is the adjustment coefficient, and \( \mu \) is the learning rate.

(11) Calculate the error of each output neuron according to the formula. If the calculation error accuracy is not greater than the given accuracy, or the maximum number of iterations is reached, the algorithm ends. Otherwise, repeat steps (5) to (11).

4. Result analysis

4.1. Analysis of Spearman results

This article first carries on relevant analysis to all the data collected by the sensor, as shown in Table 1. It can be seen from Table 1 that the relative humidity is related to evaporation, precipitation, sunshine hours, wind speed and direction, and the significance is \( P<0.05 \). And from Table 1, the degree of correlation of each variable to relative humidity (soil moisture) can be obtained.

Table 1. Spearman correlation

| Correlation coefficient | Evaporate amount | Ground temperature | Precipitation | Air pressure | relatively humidity | sunshine | Temperature | Wind speed |
|-------------------------|------------------|--------------------|---------------|--------------|---------------------|---------|-------------|-----------|
| Correlation coefficient | 1.000            | 0.633              | -0.262        | -0.415       | -0.476              | 0.342   | 0.580       | 0.319     |
| Significance (two-tailed)| 1.000            | 0.000              | 0.000         | 0.000        | 0.000               | 0.000   | 0.000       | 0.000     |
| Correlation coefficient | 0.633            | 1.000              | -0.216        | -0.784       | -0.074              | 0.376   | 0.952       | 0.024     |
| Significance (two-tailed)| 0.000            | 1.000              | 0.001         | 0.000        | 0.284               | 0.000   | 0.000       | 0.729     |
| Correlation coefficient | -0.262           | -0.216             | 1.000         | -0.021       | 0.550               | -0.501  | -0.132      | -0.009    |
Significance (two-tailed)  0.000  0.001  1.000  0.755  0.000  0.000  0.054  0.897
Air pressure  Correlation coefficient  -0.415  -0.784  -0.021  1.000  -0.132  -0.150  -0.822  ~ 0.129
Significance (two-tailed)  0.000  0.000  0.755  1.000  0.054  0.028  0.000  0.059
relatively humidity  Correlation coefficient  -0.476  -0.074  0.550  -0.132  1.000  -0.487  0.025  ~ 0.244
Significance (two-tailed)  0.000  0.284  0.000  0.054  1.000  0.000  0.720  0.000
sunshine Hours  Correlation coefficient  0.342  0.376  -0.501  -0.150  -0.487  1.000  0.256  0.029
Significance (two-tailed)  0.000  0.000  0.000  0.028  0.000  1.000  0.000  0.678
Temperature  Correlation coefficient  0.580  0.952  -0.132  -0.822  0.025  0.256  1.000  0.038
Significance (two-tailed)  0.000  0.000  0.054  0.000  0.720  0.000  1.000  0.585
Wind speed  Correlation coefficient  0.319  0.024  -0.009  -0.129  -0.244  0.029  0.038  1.000
Significance (two-tailed)  0.000  0.729  0.897  0.059  0.000  0.678  0.585  1.000
N  214  214  214  214  214  214  214  214

4.2. BAS-BP neural network model to predict moisture
In this paper, 214 groups of data are selected as the total sample. The first 174 groups of sample data are used as training samples and input into the new model for training. The remaining 40 groups are used as test samples to verify the new model. Figure 1 shows the iterative results of BAS algorithm optimization. After about 5 generations, it is close to convergence, and the result is close to the optimal fitness, and the optimal solution is 0.015. Figure 2 shows the comparison between the predicted values and the real values of BP and BAS-BP prediction models. Figure 2 shows that the change trend of the predicted value of BP and BAS-BP is the same, but the prediction error of BP model is larger, and the prediction effect is not ideal. The optimized BAS-BP prediction model has a higher degree of fit to the prediction data of soil moisture.

![Figure 1. The best fitness value of BAS optimization algorithm](image-url)
In order to verify the optimization effect of BAS on the BP algorithm, the average absolute error, average relative error, and variance of the prediction evaluation index are compared. It can be seen from Table 2 that the predictive evaluation indexes of the BP neural network model are 27.4322, 0.4023, 162.9869, and the predictive evaluation indexes of the BAS-BP neural network model are 9.1936, 0.1333, 29.5730, respectively. The above results show that the BP neural network optimized by BAS has significantly improved convergence speed and accuracy. In summary, the BAS-BP neural network model has a better prediction effect on soil moisture.

Table 2. Comparison of neural network prediction results

|                | Mean absolute error | Mean relative error | Variance   |
|----------------|--------------------|--------------------|------------|
| BP             | 27.4322            | 0.4023             | 162.9869   |
| GA-BP          | 9.7592             | 0.1442             | 48.8895    |
| BAS-BP         | 9.1936             | 0.1333             | 29.5730    |

5. Conclusion
In this paper, a BAS-BP neural network model is established by combining BAS with BP neural network to predict soil moisture. The initial weights and thresholds optimized by BAS are better than those of traditional BP neural networks. The BAS algorithm accelerates the sample training speed and convergence speed, and the prediction accuracy is significantly improved. The results show that the average absolute error and average relative error of BAS-BP are 9.1936 and 0.1333 respectively, and avoid the defect that BP neural network is easy to fall into local optimization. In this paper, the model is used to predict the change of soil moisture, which provides a new method for the prediction and evaluation of soil moisture.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (61572446) fund.

References
[1] D. Zhang, Research and Application of agricultural irrigation forecast system based on BP neural network, D. Qufu Normal University, 2019.
[2] B. Zhao, X. J. Fan, Shujuan Yi, et al, Research on dry field water-saving irrigation intelligent monitoring system, J. Journal of Chinese Agricultural Mechanization, 2016, 37 (7): 174 ~ 178.
[3] J. H. Sun, Talking about the significance of timely detection and forecast of soil moisture to
agricultural production, J. Agriculture and Technology, 2019, 39 (23): 112 – 113.

[4] G. Li, Soil moisture prediction model based on neural network, D. Beijing: Beijing University of Posts and Telecommunications, 2018.

[5] W. F. Li, Application of BP neural network in Xuchang soil moisture prediction model, J. Chinese Agricultural Science Bulletin, 2013, 29 (32): 238 - 241.

[6] L. M. Huang, S. M. Ren, Peiling Yang, et al, Simplified BP-ANN soil moisture forecast model based on default factors, J. Journal of China Agricultural University, 2013, 18 (05): 166 - 172.

[7] X. L. Hou, Y. H. Feng, Guanghui Wu, et al, Research on application of dynamic prediction model of soil moisture based on artificial neural network, J. Water Saving Irrigation, 2016 (07): 70 - 72+76.

[8] Z. S. Zhou, D. C. Wang, H. J. Sun, et al, Prediction of soil moisture based on GA-BP algorithm, J. Journal of Heilongjiang Ba Yi Land Reclamation University, 2009, 21 (05): 73 - 75+104.

[9] M. S. Li, Z. Liu, Application of soil water balance model in soil moisture forecast, J. Northeast Water Resources and Hydropower, 2005, 23 (1): 49 - 56.

[10] X. L. Wang, G. Shi, M. Y. Zhang, et al, Research on the correlation of computer science and technology professional curriculum system based on pearson coefficient, J. Wireless Internet Technology, 2019, 16 (21): 114 - 115.

[11] H. F. Ai, H. Pan, Application of BP network in prediction of agricultural machinery in Jilin Province, J. Journal of Chinese Agricultural Mechanization, 2016, 37 (8): 208 - 211.

[12] X. Y. Chen, L. Li, Xiaojing Fan, et al, Farmland pest prediction algorithm based on neural network and evidence theory, J. Journal of Chinese Agricultural Mechanization, 2019, 40 (2): 151 – 156.

[13] L. Y. Wang, S. G. Guo, Water absorption index prediction for corn gluten meal based on BP neural network, J. Journal of Chinese Agricultural Mechanization, 2013, 34 (6): 125 - 128.

[14] Y. J. Xiang, S. Li, Beetle antennae search algorithm for optimization problems, J. International Journal of Robotics and Control, 2018, 1 (1).

[15] X. Y. Jiang, S. Li, Beetle antennae search without parameter tuning (BAS-WPT) for multi-objective optimization, J. Preprint, 1.