Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 10, No. 2, 2024

Prediction of Sunspot Activity Based on Differential Evolution Algorithm and BP Neural Network Model

Ziyang Yu [1], Xinpeng Yu [2], Yinuo Liu [3], Yue Li [4], Yuyan Zeng [1], Yanqing Liu [1, *]

[1] School of Information Engineering, Guilin Institute of Information Technology, Guilin, Guangxi, China
[2] School of Electronic Engineering, Guilin Institute of Information Technology, Guilin, Guangxi, China
[3] School of Foreign Trade and Foreign Languages, Guilin Institute of Information Technology, Guilin, Guangxi, China
[4] School of Mechanical and Electrical Engineering, Guilin Institute of Information Technology, Guilin, Guangxi, China

*** Corresponding author:** Yanqing Liu (Email: 837104905@qq.com)

Abstract: In this research report, the prediction method of sunspot activity is deeply discussed, and a variety of statistical and data science models are used to make comprehensive prediction. Research focuses include the start time and duration of the solar maximum in the solar cycle, as well as the prediction of the number and area of sunspots. By capturing the periodicity of solar activity, Pearson correlation coefficient is used to analyze the relationship between the maximum solar activity and the number of sunspots, and the adaptive multivariate nonlinear regression-BP neural network model and particle swarm optimization BP neural network are combined to make high-precision prediction. The research results provide a scientific basis for the prediction of space weather, ionospheric state and communication system reliability.

Keywords: Neural Network; Pearson Correlation Coefficient; Differential Evolution Algorithm; Particle Swarm Optimization.

1. Introduction

This research report aims to analyze and predict the key features of sunspot activity in depth. Sunspots are temporary spots on the photosphere whose number and area changes are closely related to the periodicity of solar activity [1]. To better understand and predict sunspot activity, this study used a variety of statistical and data science models. Specifically, we used Pearson correlation coefficient analysis to explore the relationship between solar maximum duration and sunspot number [2], and built a prediction system [3] based on the adaptive multivariate nonlinear regression-BP neural network model to predict the start time and duration of the solar maximum in the next solar cycle. In addition, in order to more accurately predict the changing trend of sunspot number and area, we introduced the time series prediction model of BP neural network based on particle swarm optimization [4]. This model combines the prediction ability of BP neural network and the optimization effect of PSO algorithm [5], and can more accurately reflect the dynamic changes of sunspot activity. Through the application of these comprehensive methods, it is expected to provide a more accurate and reliable basis for the prediction of sunspot activity, and further promote the development of solar physics, space weather prediction, communication technology and other related fields.

2. Prediction of the Maximum Solar Value in the Solar Activity Cycle

In order to predict the start time and duration of the solar maximum in the next solar activity cycle, Pearson's correlation coefficient detects that there is a correlation between the maximum and the solar black number. It can be seen that the correlation is strong through the thermal map. Therefore, a single linear regression is established to obtain the negative correlation between them before. Based on adaptive multivariate nonlinear regression -BP neural network model for solar activity prediction, first establish the relationship between time and the number of sunspots model, and then based on the differential evolution algorithm model for solving, and then use the differential evolution algorithm model and BP neural network model for mixed model to solve, calculate the maximum for back generation, to achieve the next cycle prediction.

First of all, the official data was merged and cleaned, and the historical data of the sunspot area was collected by means of python data crawler and ESA.

2.1. Pearson's Correlation Coefficient

By collecting and processing the start time and duration of the maximum value of each phase, as shown in Figure 1:

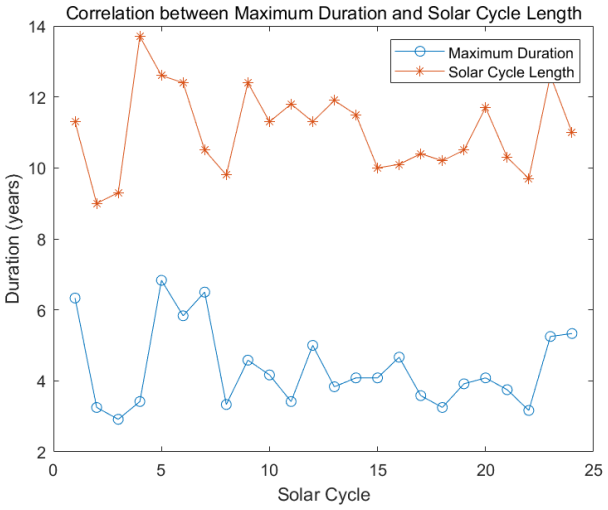

Figure 1. The start time and duration of the maximum value

The figure shows the relationship between the solar cycle length and the maximum duration. Through comparison, it can be seen that when the solar cycle length is shorter, the maximum duration is usually shorter. When the solar cycle length is longer, the maximum duration may reach a relatively high value. It is found that there is a strong relationship

between the start time and the duration of the maximum of each period. Therefore, in order to highlight the significance of the relationship, Pearson's correlation coefficient is introduced.

The Pearson correlation coefficient applies to linear correlation relationships. The degree of correlation is related to the p-value, which is its correlation coefficient. The closer the correlation coefficient is to 1, the stronger the correlation between the two factors is said to be; On the contrary, the closer the correlation coefficient goes to 0, the weaker the correlation.

The Pearson correlation coefficient is often used to measure the degree of correlation (linear correlation) between two variables X and Y, with values between -1 and 1. Introduce covariance and variance here:

The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} \quad (1)$$

Establish a heat map to further observe its significance.

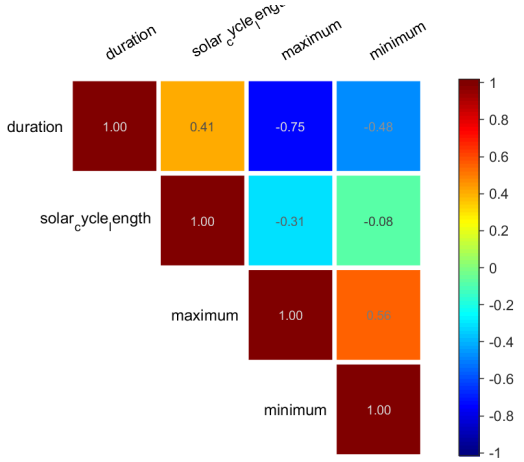

Figure 2. Duration and solar cycle duration heat map

As can be seen in Figure 2: duration has a moderate positive correlation with solar cycle duration, a strong negative correlation with the maximum value of the number of black characters, and a moderate negative correlation with the minimum value of the number of black characters.

2.2. Unary Linear Regression

It can be seen from the above that the maximum number of sunspots shows a strong negative correlation with the duration, so the correlation expression is based on the single linear regression.

Given that $p<0.05$, the regression model $y=6.9871-0.0147x$ was established.

2.3. Establishment of the Relationship Model Between Time and Sunspot Number

We need to build a regression analysis model to quantify the relationship between time and the number of sunspots. In regression analysis, we usually use independent variables (time) and dependent variables (number of sunspots) to build models. By fitting this data, we can get a regression equation that describes the relationship between time and the number of sunspots. This equation can be used to predict future trends in sunspot numbers. The relationship can be expressed as:

$$y_i = f\left(x_1^i, x_2^i, \cdots, x_j^i, \theta_1, \theta_2, \cdots, \theta_p\right) + \sigma_i \varepsilon \left(i = 1, 2, \cdots, n\right) \quad (2)$$

Where, y is the true value; i represents group i data.

$f(x_1, x_2, \cdots, x_j, \theta_1, \theta_2, \cdots, \theta_p)$ is a multivariate nonlinear function, representing the deterministic part.

$x_1, x_2, \cdots, x_j$ is the independent variable; $\theta_1, \theta_2, \cdots, \theta_p$ is an unknown model parameter of a multivariate nonlinear function.

$\sigma_i \varepsilon$ is the random part, ε is a random variable that follows the N (0,1) distribution.

σ_i is the random distribution standard deviation of group i data.

At present, the selection of nonlinear regression models mainly relies on empirical or experimental methods [6].

In order to solve the problem of greater error in the empirical method, we can use the experimental method to determine the regression model suitable for describing the relationship between time and the number of sunspots.

When choosing a regression model, we can find the most suitable one by comparing different nonlinear regression models.

Through this series of experimental steps, we can more accurately determine the regression model that is suitable to describe the relationship between time and the number of sunspots, and obtain more accurate results. Therefore, in this paper, the regression model suitable for this relationship is determined by experimental method as follows:

$$y = x1 * Sin\left(x2 * x + x3\right) + x4 \quad (3)$$

2.4. Model Solving Based on Differential Evolution Algorithm

We are faced with a multi-parameter optimization problem in which several parameters need to be determined. To solve this problem, we consider using differential evolution algorithm to optimize the solution.

In practical applications, we combine the differential evolution algorithm with the neural network model to further improve the prediction accuracy and stability. By adjusting the parameters of the differential evolution algorithm and setting the fitness function, we can better control the optimization process and obtain better prediction results.

Therefore, we choose differential evolution algorithm, which can better solve the global optimal problem, to optimize and solve the parameters [7].

The following are the solving steps:

Through differential evolution algorithm optimization, we find the optimization results of seven parameters. The results are:

$$\begin{aligned} x1 &= 26.9884001822349 \\ x2 &= 0.0444929670892593 \\ x3 &= -1.90076018991413 \\ x4 &= 81.9846295969476 \end{aligned} \quad (4)$$

Through the analysis, we can draw the following conclusions: 1. Differential evolution algorithm can effectively deal with regression problems and get better fitting results. 2. Filtering technology can effectively remove the noise in the data and improve the accuracy and reliability of data. 3. By comparing the actual value with the filtered value, it can be seen that the filtering technology can smooth the data

better, so as to better reflect the overall trend and rule of the data.

2.5. BP Neural Network Prediction Model

BP neural network is a multi-layer feedforward algorithm, which is composed of input layer, hidden layer and output layer. There are working signals and error signals propagating between layers [8].

The operation principle of BP neural network is as follows:

Remember the training set as $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, $x_i \in R^d, y_i \in R^l$ and the output as $\hat{y}_k = (\hat{y}_1^k, \hat{y}_2^k, \ldots, \hat{y}_l^k)$ .then: $\hat{y}_j^k = f(\beta_j - \theta_j)$ $\beta_j = \sum_{i=1}^{n} w_{ij} x_{ij}$.

Where, w_{ij} is the connection weight of the i th neuron to the j th output. When the network is recorded on $(\boldsymbol{x}_k, \boldsymbol{y}_k)$, the error is:

$$E_k = \frac{1}{2}\sum_{j=1}^{l}\left(\hat{y}_j^k - y_j^k\right)^2 \tag{5}$$

When the neural network completes the forward calculation, the predicted value is subtracted from the actual value to get the error value, and then the weight threshold of the neural network is adjusted by backpropagation. The iterative updating formula of w and θ is as follows:

$$\begin{aligned} \Delta\omega_{hj} &= \eta g_j b_h \\ \Delta\theta_j &= -\eta g_j \\ g_j &= -\frac{\partial E_k}{\partial \hat{y}_j^k} \bullet \frac{\partial \hat{y}_j^k}{\partial \beta_j} = -\left(\hat{y}_j^k - y_j^k\right) \bullet (\hat{y}_j^k)' \end{aligned} \tag{6}$$

Where, b_h is the input data of the neuron. Based on this, the neural network constantly adjusts the weights and thresholds in its training process, so that the prediction error of the neural network is constantly approaching 0.

2.6. Hybrid Model Solution Results

For each period, the observed value is the data that actually occurred, the model predicted value is the prediction result based on the BP neural network prediction model, and the fit value of the autoregressive moving average model is the fit result based on the differential evolution algorithm. By comparing the values of these three labels, we can assess the model's predictive power and fitting effect.

From the analysis, we can see that the maximum value occurs in January 2036, and the corresponding sunspot quantity is 154.3.

Therefore, back can be performed, and the result is shown in Table 1:

Table 1. Prediction result

Solar cycle	Start time	End time	Maximum duration	Solar cycle duration
23	1996-08	2001-11	5.25	12.6
24	2008-12	2014-04	5.333333333	11
25	2019-12	2025-5	5.5171	10.839
26	2031-9	2037-3	5.368	10.8

In order to predict the start time and duration of the solar maximum in the next solar activity cycle, Pearson's correlation coefficient detects that there is a correlation between the maximum and the solar blackness number. Therefore, a single linear regression is established to obtain the negative correlation between them before. Based on adaptive multivariate nonlinear regression -BP neural network model for solar activity prediction, first establish the relationship between time and the number of sunspots model, and then based on the differential evolution algorithm model for solving, and then use the differential evolution algorithm model and BP neural network model for mixed model to solve, calculate the maximum for back generation, to achieve the next cycle prediction.

3. Prediction of the Number and Area of Sunspots

To predict the number and area of sunspots in the current and next solar cycles. We introduce FLY to analyze periodicity, and combine the time series prediction model based on BP neural network of particle swarm optimization to build a time prediction trend graph.

3.1. FFT Analyzes Periodic Data

Sunspot forecasts are usually derived from monthly averages. For this, monthly smoothing of sunspot numbers is performed to plot the monthly average of sunspots over the period approximately 1700-2000. To get a more detailed view of the periodic nature of sunspot activity, 50 years of data will be plotted, as shown in Figure 3.

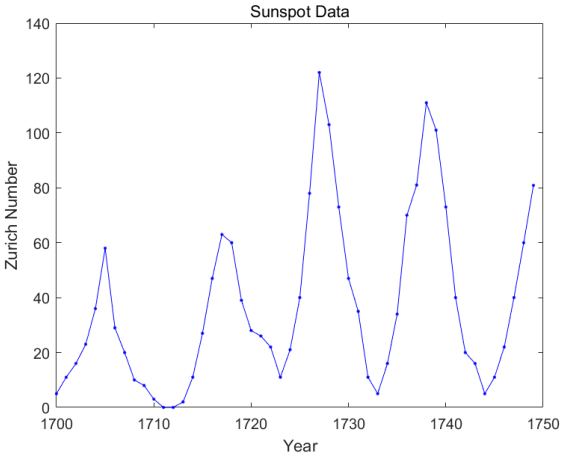

Figure 3. Monthly average value of sunspot number

Therefore, the number and area of sunspots can be analyzed periodically.

3.2. Establishment of PSO-BP Prediction Model

During the establishment of BP neural network, the random setting of connection weights will lead to errors in the prediction results, and gradient descent training has the shortcomings of slow speed and local minimum, so it is difficult to achieve the global optimal training of neural network. Particle swarm optimization algorithm (PSO) is used to optimize BP neural network, and the prediction accuracy and generalization ability are improved. The construction process is as follows:

(1) Data normalization, establish BP neural network, determine the topology and initialize the weight and threshold of the network;

(2) Initialize PSO parameters, such as the maximum number of iterations, population size, individual learning factor, social learning factor, inertia weight, etc.;

(3) Initialize the population position of PSO, and calculate the number of variable elements to be optimized according to the BP neural network structure;

(4) PSO optimization, fitness function is set as the mean square error predicted by BP network, cycle the PSO optimization process, constantly update the position of the optimal particle until the maximum number of iterations, and terminate the PSO algorithm;

(5) The optimal weight threshold parameter after optimization of the PSO algorithm is given to the BP neural network, that is, the optimal PSO-BP model is output, and PSO-BP is used for training and prediction, and compared with the BP network before optimization.

3.3. Model Construction

In the above discussion of "monthly average area of sunspot number", it can be seen that there are 1794 sets of data through data processing. In the time series regression prediction, we classified 1000 sets of data as training set, 794 sets of data as test set, and iterated 15 sets of data to predict the 16th data, and so on. It was found that the predicted data and frame number data had a good fit.

Therefore, we added two more groups of data numbers to be predicted in the test set, and we can know that the area of sunspots in the current solar cycle is 1075.56 through the prediction.

This gives us a total sunspot area of 1068.95 in the next cycle.

In the above discussion of "total number of smoothed sunspots per month", we can see a total of 3286 sets of data through data processing. In the time series regression prediction, we classified 2000 sets of data into the training set, 1286 sets of data into the test set, and iterated 15 sets of data to predict the 16th data, and so on. It was found that the predicted data and frame number data had a good fit.

Therefore, we add two more groups of data numbers to be predicted in the test set, and we can know that the number of sunspots in the current solar cycle is 133.2 through prediction.

This gives us a total of 156.3 sunspots in the next cycle.

When optimizing BP neural network for time series prediction, the weight and threshold of BP neural network can be represented as the position of the particle, and mean square error or other appropriate evaluation index can be used as fitness function. By constantly updating the position and velocity of the particles, the prediction error of BP neural network is gradually reduced, and better prediction results are obtained.

The PSO is initialized to a group of random particles (random solutions). The optimal solution is then found through iteration. In each iteration, the particle updates itself by tracking two "extreme values "(pbest, gbest). After finding these two best values, the particle updates its speed and position by the formula.

3.4. Model Reliability

Predict the accuracy rate through various error indexes of BP and PSO-BP.

Table 2. Error indexes of BP and PSO-BP

	Mean absolute error mae	Mean square error mse	Mean square error root rmse	Mean absolute percentage error mape	BP forecast Accuracy	PSO-BP prediction accuracy
BP Neural Network	3.9539	22.2639	4.7185	31.7068%	/	/
pso optimizes BP neural network	1.2073	2.4017	1.5498	8.2818%	68.2932%	91.7182%

As can be seen from Table 2, the prediction accuracy of the model can be indirectly predicted by the various error metrics of BP and PSO-BP. We found that: comparing the error metrics of BP and PSO-BP, including mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE), etc. the error metrics of PSO-BP are significantly lower than those of BP.

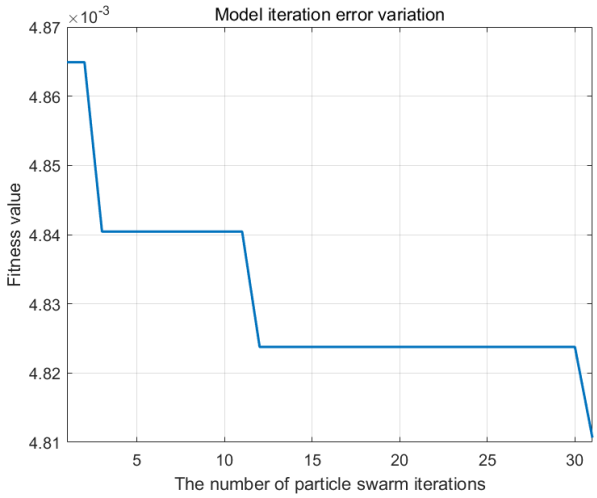

Figure 4. Relationship between model iteration error and particle number

Regression graphs of BP and PSO-BP models, using the total number of predicted sunspots as an example.

The Figure 4 shows the relationship between model iteration error and particle number, with the horizontal coordinate representing particle number and the vertical coordinate representing model iteration error. As can be seen from the figure, as the number of particles increases, the iteration errors of the model show a trend of decreasing first and then increasing. The position and velocity changes of particle swarm during PSO-BP optimization were observed. It is found that the particle swarm can find a better solution, and the position and velocity change trend is relatively stable during the iteration process, so it can be considered that the prediction accuracy of the model may be higher.

It can be seen from the above that the model has high accuracy and strong reliability.

In order to predict the number and area of sunspots in the current and next solar cycle. We introduce FLY to analyze periodicity and find that the number and area of sunspots have certain periodic changes. Combined with the time series prediction model based on BP neural network of particle swarm optimization, a time prediction trend graph is established. Analyze and process "monthly average sunspot area" and "smooth monthly sunspot number", and use 15 sets of data as an iterative unit to predict the next set of data for iterative training. The two data sets are divided into training set and test set in turn. According to the fitting trend of the

training set, it is found that the real value and the predicted value have a high coincidence degree. Therefore, the number and area of sunspots in the current and next solar cycle were obtained by fitting the test set and particle swarm optimization was introduced to optimize the algorithm. In order to test the reliability of the model, the prediction accuracy of BP neural network was found to be 68.2932% through the prediction and inspection of BP neural network. And the accuracy of BP neural network time series prediction model based on particle swarm optimization is 91.7182%, and its model has certain reliability.

4. Conclusion

In this study, Pearson correlation coefficient analysis, adaptive multivariate nonlinear regression-BP neural network model and particle swarm optimization BP neural network were used to construct a comprehensive prediction model of sunspot activity. The prediction results show that the model can accurately predict the start and end time of the solar cycle, the start time and duration of the solar maximum in the solar cycle, as well as the number and area of sunspots. In addition, the reliability of the time series prediction model based on particle swarm optimization BP neural network is verified by comparing it with BP neural network model. This study not only provides a scientific basis for the prediction of sunspot activity, but also provides a reference for the prediction of other data with periodic changes.

References

[1] LIKejun SUTongwei LIANGHongfei.Periodicity of sunspot activity in the modern solar cycles[J]. Chinese Science Bulletin, 2004, 49 (21) : 2247-2252.

[2] Abdisa G. Dufera,Tiantian Liu,Jin Xu.Regression models of Pearson correlation coefficient[J].Statistical Theory and The Related Fields, 2023, 7 (02) : 97-106. The DOI: 10.1080/ 2475 4269. 2023.2164970.

[3] Daniel Okoh, Loretta Onuorah,Babatunde Rabiu, et al.An application of artificial intelligence for investigating the effect of COVID-19 lockdown on three-dimensional Temperature variation in equatorial Africa [J]. J team Frontiers, 2022, 13 (02): 52-61. DOI: 10.1016 / j.g sf. 2021.101318.

[4] Qiang Li,Miao Wan,ShuGuang Zeng, et al.Predicting the 25th solar cycle using deep learning methods based on sunspot area data[J].Research in Astronomy and Astrophysics, 2021,21 (07):290-298. (in Chinese) DOI:10.1088/1674-4527/21/7/184.

[5] LIU Shijun, YU Xiaoding & CHEN Yongyi Chinese Meteorological Administration Training Centre, Beijing , China Correspondence should be addressed to Liu Shijun.Prediction of solar cycle based on the invariant [J].Chinese Science Bulletin,2003,48(23):2568-2571.

[6] Yibin Yao,Bao Zhang,Chaoqian Xu, et al.Analysis of the global T_m–T_s correlation and establishment of the latitude-related linear model[J].Chinese Science Bulletin,2014, 59(19): 2340-2347.

[7] Wei Qian, Yanmin Wu,Bo Shen.Novel Adaptive Memory Event-Triggered-Based Fuzzy Robust Control for Nonlinear Networked Systems via the Differential Evolution Algorithm [J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11 (8) : 1836-1848. The DOI: 10.1109 / JAS. 2024.124419.

[8] Tang, Lin, Xu, Zhi-Pei,He, Tian-Long, et al.Sensitivity analysis of rotation frame stability based on BP neural network[J].Chinese Journal of Engineering The Design, 2018, 25 (5) : 576-582. The DOI: 10.3785 / j.i SSN. 1006-754 - x. 2018.05.012.