

Research on Pepper Disease Diagnosis Model Based on Genetic Neural network

Qingfeng WEI^{1,2}, Sufen SUN^{1,2}, ChangShou LUO^{1,2}

¹Institute of Agricultural Information and Economics, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

²Beijing Research Center of Engineering Technology on Rural Distance Information Service, Beijing, China

Abstract. To explore the effective method of the pepper diseases, using the method of combining the genetic algorithm and BP neural network, the establish neural network diagnostic model for pepper diseases is established to improve the model convergence rate and accuracy. The experimental results show that the model is convergent effectively and the diagnostic accuracy is improved by 12%.

1 Introduction

Pepper is rich in vitamins. It can enhance the physical strength, dispel cold and clear damp, and has been one of the daily necessities of vegetable species. Pepper industry is one of the most important vegetable industries in China. Pepper disease is an important factor restricting high quality production. If the disease can not be diagnosed effectively, the best prevention and control time will be delayed, the quality of agricultural products will be reduced and the losses will be greater.

Recent years, BP neural network is widely used in diagnosis, which provides a new way for the diagnosis of plant diseases. But throughout the current study, there are major problems: Firstly, the traditional method (additional momentum and adaptive learning rate method) for performing neural network performance optimization is an iterative search process which the optimal solution is obtained from an initial point in the solution space. The search information provided by a single search point is not much, so the search efficiency is not high in the intention to obtain the global optimal solution. Secondly, these algorithms often use deterministic search methods. Namely, one search point to another, there is a definite transfer method and transfer relationship. This kind of certainty often makes the search process fall into the local optimal solution and stagnates, and even makes it impossible to reach the optimal solution forever.

In view of the existing problems in the application of the model, the genetic algorithm is used to optimize the weights of the neural network in order to improve the model and improve the accuracy of diagnosis.

2 BP neural network model construction

2.1.1 Input matrix construction

Pepper disease symptoms are mainly reflected in the whole plant of seedling stage, and 5 parts of roots, stems, vines, leaves, flowers, fruit. Therefore, the input vectors are divided into 6 units. The contribution of different symptoms to disease diagnosis is different. Some of the obvious symptoms are often the important basis for the identification of certain diseases. Invite 5 experts, according to the contribution of disease symptoms to disease diagnosis, making in the root, stems, leaves, flowers, fruit, seedlings, and take the average value to build input vector.

Table 1 symptom types and scores

Type	Score	Explain
Typical symptom	1	The type of disease can be determined by this type of symptom.
Main symptom	0.7	Through the comprehensive expression of several symptoms, can determine the type of disease.
General symptoms	0.3	The symptoms only serve as an auxiliary judgment, and can not determine the disease types.

2.1.2 Hidden layer neuron determination

The hidden layer is the value map that is input to output. The output formula of hidden layer neuron is as follows:

$$y = f\left(\sum_i w_{ji}x_i - \theta_j\right)$$

Among them, “y” is the hidden layer node value, “w_{ji}” is the weight between the input point and the hidden layer neuron node, and “θ” is the threshold value.

The activation function is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The number of neurons in the hidden layer is determined as follows:

$$l = \sqrt{n + m} + a \quad 0 < a < 10$$

Among them, "l" is the number of neurons in the hidden layer, "n" is the input layer neuron number, "m" is the output layer neuron number, and "a" is the constant value between 0 and 10.

2.1.3 Output matrix construction

The output vector is a matrix of n*n. It constructed by binary encoding of "1 in n". Among them, n is the length of coding, that is, the total number of diseases. Each set of codes represents a kind of disease. Only 1 of them were 1, and the other N-1 was 0. The location of 1 is different. The types of diseases are different. The closer the maximum of the element in the vector to 1, the greater the likelihood that the disease will occur, and the closer it is to 0, the smaller the likelihood of the disease.

3 Optimization of initial weights of neural networks based on Improved Genetic Algorithm

A genetic algorithm is based on real number coded generally. The range of cross coefficient is [0, 1]. When it is used to optimize the weights of the neural network model, it will cause the chromosome individuals to contract or diverge in the search space. Therefore, the optimization ability is reduced, which is not conducive to global optimization. For variability, if the variability is too low, the search is too narrow and the evolution rate is slow. If the amount of variability is too large, the search will oscillate near the target point, and the optimization goal can not be achieved. To those problems, the improvement method is: To the parent chromosome individual in an initially uniform distribution, increase the dispersion degree of the population after the crossover operation to generate new search space. At the same time, the offspring are distributed evenly in the search space. In the search space, neither contraction nor divergence, and ensure that individuals with excellent patterns are not completely destroyed by crossing operation, so as to improve the global optimization ability. For mutation operators, when the evolutionary process is smooth, the rate of variability is maintained at a low level. When evolution is not successful, the search range is increased by increasing mutation rates. When the evolutionary algebra reaches a set threshold, it is assumed that the population has reached an early maturity. Forcing the population out of an early maturing state by forced mutation. The step is as followed:

Step1: The chromosome gene coding is based on the structure of the neural network model and the solving problem.

Step2: Randomly generated initial and uniform chromosomes population.

Step3: According to the problem, calculate the fitness values of individuals in a population. Use training sample to train the neural network represented by the individual in the group, calculate the learning error of each individual, and then determine the fitness value. Learning error E is calculated by the lower form:

$$E = \sum_{k=1}^n E_k, \quad E_k = \sum_{j=1}^q (y_j^k - C_j^k)^2 / 2$$

Among them, "n" is the training sample number. "q" is the number of output units. The represents the difference between the actual output and expected output of the k sample training the j output unit. "fs=1/E" is fitness function. It ensures that the smaller the error, the greater the fitness value.

Step4: Genetic manipulation of individuals in a population, including selection, crossover, and variation.

Selection: Roulette method is used to select the operation.

Cross: If x1, X2 are the parent individuals, and uniformly distributed random numbers in V=[xmin, xmax] , then, the uniformly distributed random numbers Z1 and Z2 on the interval [xmin, xmax] after cross is calculated by the following formula:

$$y1=ax1+(1-a)x2, \quad y2=ax2+(1-a)x1 \\ z1=MOD(y1,V), z2=MOD(y2,V)$$

Z1 and Z2 are the offspring after crossover, in which, a is integer and MOD is modulo operator.

Variation: the rate of variation Pm is defined by the following:

$$Pm=0.001+NG*cof$$

Among them, Pm represents the rate of variability of the current algebra. NG is an algebra that has not evolved since the last evolution. COF is the coefficient that determines the threshold of chromosome variation (100% variation). The variance quantity Var is defined as:

$$var = rand * w * dyna$$

In the formula, rand is the random number between [0, 1]. "w" is a fixed value within the value range of the weight. "dyna" is the dynamic parameters of Var. Initial dyna= 1.0. If "counter" > "nochange", "dyna"="dyna0.1" and "counter"= 0. Among them, "counter" is an algebra that has not evolved since the last evolution until the present generation. "nochange" is a constant, is the threshold to decide whether to change the "dyna".

Step5. Calculate fitness value again.

Step6 If the stopping search condition is satisfied, the iteration stops and the output solution is the optimal solution. Otherwise, the steering step D. At the same time, In order to ensure that individuals with the best patterns are not damaged by genetic manipulation, the optimal individual retention strategy is adopted. That is, the individuals with the greatest current fitness value are genetically transmitted to the next generation without genetic manipulation. The improved genetic algorithm guarantees the stability and convergence of the optimized neural network model, and makes the improved genetic algorithm always have higher optimization efficiency in the optimization of the neural network model.

4 Model test

Take 15 kinds of diseases as an example, such as scab, powdery mildew, leaf blight etc. 6 dimensional input vectors and 15 dimensional output vectors are constructed by the above method. In Matlab, the hidden layer units is 10, the target error is 0.0001, and the loop 1000 times, and the Levenberg-Marquardt algorithm is used to train, and statistical accuracy of diagnosis.

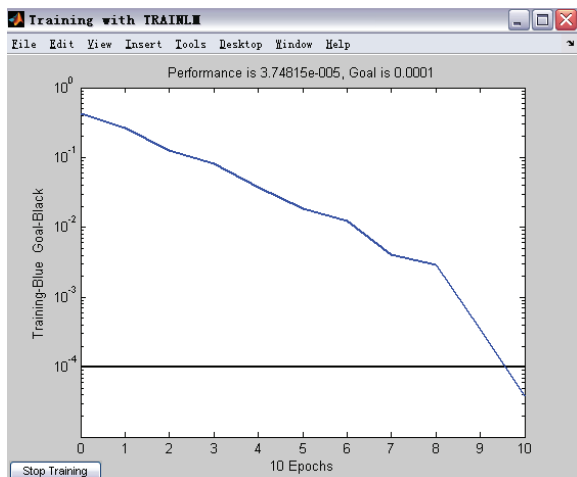


Fig.1 Test data convergence

Table 7 The test results of different models

Models	Diagnostic accuracy		Average accuracy
	Laboratory test sample data	Field application data	
BP neural network	73.70%	58.20%	66%
Genetic neural network	82.60%	73.40%	78%

The accuracy of this model is compared with BP neural network. The test data consists of two types: Laboratory test sample data, farmers generated data according to field symptoms. Test statistics show that, Whether laboratory data or household data, the diagnosis method based on genetic neural network is better than the BP neural network in correct rate. It shows that the improved genetic algorithm is more effective in improving the accuracy of the model. And the result of laboratory data test is better than that of peasant data. The reason is that, laboratory data reflect the relationship between a disease and a class of symptoms. In the field of actual production, there are many diseases at the same time, and a kind of symptoms corresponding to many diseases. When this happens, the diagnostic accuracy is influenced. It also shows that efforts in this regard can further to improve the usefulness of the model.

5 Conclusion

In this paper, the crossover mutation operator of genetic algorithm is improved, the improved genetic algorithm is used to optimize the weights of the neural network, and then, a pepper disease diagnosis model based on Genetic Neural Network is established. Tests were carried out on

15 kinds of pepper diseases, the accuracy of the model is improved by 12% compared with the general BP neural network model.

Acknowledgment

The research work was supported by Beijing science and technology plan project: Beijing science and technology correspondent resources service platform construction and demonstration, the fund for agriculture of Beijing municipal commission of rural affair: The application of APP and "zixuntong" products in the management and service of general agricultural technicians, the young scientist fund of BAAFS: Research on agricultural scientific technical online counseling system based on the knowledge map (NO.QNJJ201534), the innovation fund of BAAFS: Agricultural water-saving technology training and counseling service demonstration project (NO.KJCX20151409). The correspondence author: Sufen SUN(Sunsf@agri.ac.cn).

References

1. J.C. LAI, B. MING, S.K. LI ect. Agricultural Sciences in China Vol.8(2010), p.1222
2. Mansin gh G, Reichgelt H, Bryson K-MO. CPEST: Expert Systems with Applications Vol.32(2007),p. 185.
3. Ghosh I, Samanta R K. Teapest: Applied Engineering in Agriculture Vol.19(2003), p. 620
4. Y.F. CUI, S.P. WANG:Journal of shanghai jiaotong university Vol.29(2011),p.80
5. Li-feil YOU, Xiao-qinZHONG, Chang-yuan LI. Study of the Evaluation Method of Management Diagnosis Scheme Based on Modified Artificial Neural Network. Journal of Guangzhou Maritime College. Vol 2(2017), p 35-38.
6. ZHU Lizhong, CHENG Nan. Fault Diagnosis of SCADA System for Artificial Neural Network. Transactions of Shenyang Ligong University. Vol 2(2017), p102-106.
7. ZHANG Wen-jing, ZHANG Lei, LIU Ming.Fault Diagnosis of Wind Turbine Gearbox Based on Wavelet Transform and BP Neural Network. Techniques of Automation and Applications. Vol 6(2017), p1-4.
8. Lv Shenglin, Jiang Hongjun, Yu Sidong. Optimization Analysis of Automobile Plastic Part CAE Based on BP Neural Network and Genetic Algorithm. Modern Plastics Processing and Applications. Vol 2(2017), p 55-59.
9. SUN Xianping YANG Bing LIU Qiangqiang etc. Warm Extrusion Die Wear Optimization Design Based on BP Neural Network and Genetic Algorithm. Lubrication Engineering. Vol 4(2017), p 84-88.
10. WANG Li-min,QIAO Ling-ling,WEI Lin-jing. Optimal convolutional neural networks learning method combined with genetic algorithm. Computer Engineering and Design. Vol 7(2017), p1945-1950.