

# Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm

Changshou Luo<sup>1</sup>, Qingfeng Wei<sup>1</sup>, Liying Zhou<sup>2\*</sup>, Junfeng Zhang<sup>1</sup>, Sufen Sun<sup>1</sup>

<sup>1</sup>:Institute of Information on Science and Technology of Agriculture, Beijing Academy of Agriculture and Forestry Sciences,Beijing,100097, PR China; <sup>2</sup>: China Agricultural University Library, Beijing 100094, PR China

**Abstract.** In this paper, the theory and construction methods of four models are presented for predicting the vegetable market price, which are BP neural network model, the neural network model based on genetic algorithm, RBF neural network model and an integrated prediction model based on the three models above. The four models are used to predict the Lentinus edodes price for Beijing Xinfadi wholesale market. A total of 84 records collected between 2003 and 2009 were fed into the four models for training and testing. In summary, the predicting ability of BP neural network model is the worst. The neural network model based on genetic algorithm was generally more accurate than RBF neural network model. The integrated prediction model has the best results.

**Keywords:** genetic algorithm, neural network, prediction, vegetables price

## 1 Introduction

There is a fashion to say: " food is the god of people, Vegetables is the half of it" .Vegetables industry plays an important role in providing abundant fresh agricultural products. It is also an important source of peasant's income. Vegetables price is unstable and change fast. The prediction is difficult. All those became the main obstacle factors to promoting the sustained and steady development of the vegetable production. Use the scientific method to excavate the change law of the vegetables price. And forecast the trend immediately and accurately. All those works have great significance to vegetable production, government regulation and vegetable industry stabilization. In the current research, non-linear prediction methods such as neural network and genetic algorithm are used widely, and also make certain results. For the complexity of vegetable price forecasting, BP nerve network model, the neural network model based on genetic algorithm, RBF neural network model are set up separately. And on the basis of them, an

---

\* Corresponding author. Tel.:+86 10 62731045  
*E-mail address: zhouly@cau.edu.cn*

integrated prediction model is established to provide the price of agricultural production with a reference of accurate prediction.

## **2 Materials**

Market prices of agricultural products are affected by many factors such as climate, supply and demand etc. The prediction is more complicated than commercial products. It is very difficult to collect the data of impacting factor accurately and timely. Therefore, in this study, take the vegetable price as experimental materials. The daily price of Lentinus edodes is obtained in Beijing Xinfadi wholesale market from 2003 to 2009. After it was weighted, the monthly price is gain to prepare for forecasting.

## **3 Experimental methods**

### **3.1 Construction of BP neural network model**

Artificial neural network model (abbreviations is ANN) can deal with some problems that background information is not clear and inference rule is uncertain. In theory, it can achieve arbitrary nonlinear mapping from input to output. So, it is widely used in complex nonlinear prediction, such as electricity prices and stock prices[1,2], and crop pests[3,4]. BP neural network (abbreviations is BPNN) is a kind of ANN which weight adjustment use back propagation learning algorithm. It has three layers which contain input layer, output layer and hidden layer. Connections are between upper neurons and lower neurons. But there is no link within the same layer.

In this paper, take vegetable price data of a certain phase as input, the others as output to construct the BPNN.

The number of neurons in hidden layer is important for forecasting. For the own characteristics of neural network, there is not a formula to determine the number of neurons in hidden layer. Practical experience shows that, with the increase of the neurons, the amount of computation is large, the simulation is better, but the generalization ability is poor. For complex matters, too few neurons can not reveal the variation. Generally, the number of neurons is more appropriate between 5 and 30. The exact number needs designer's experience and experiments to determine.

The conversion from input layer to hidden layer is realized by the function of `tansig()`, and from hidden layer to output layer is realized by the function of `purelin()`. For the BPNN has the shortcoming which is easy to local minimums, the training algorithm of Levenberg-Marquardt is used to improve the optimization for efficiency.

### 3.2 Construction of neural network model based on genetic algorithm

Genetic algorithm (abbreviations is GA) was proposed by Holland professor in 1975. It is an optimal method based on natural selection and evolution in high dimension space. For the ability of global optimization, it was widely used in model optimization. Complement of GA, the BP neural network can solve the problems of falling in local minimums. Lots of literature used the method to forecast, such as soil salt distribution, electricity price, and so on [5-7].

Taking the monthly market price of Lentinus edodes which is from 2003 to 2007 as experimental data, the neural network based on GA is constructed. Its structure is similar to BPNN. The main process of using GA for optimization of neural network model is as follow:

(1) Gene Encoding. According to the BPNN, gain its weight number. Every weight is on behalf of a gene. All of them structure a chromosome.

(2) Initial chromosome group generation. Choose the number of chromosome in initial population. To each chromosome, generate weights randomly in the given range to construct the initial group.

(3) Individual fitness computation. Use training samples to train the individual chromosome which is on behalf of an ANN, and then, calculate the individual learning error E. The formula is as follow:

$$E = \sum_{i=1}^n E_i, \quad E_i = \sum_{l=1}^m (y_l^i - c_l^i)^2 / 2 \quad (1)$$

Here, n is the number of training sample, m is the number of output unit.  $y_l^i - c_l^i$  is the difference between actual value and expected value of l-output when it takes i-sample to train. Fitness function fs is as bellow:

$$fs = 1 / E \quad (2)$$

(4) Selecting operation. Select the individual taking roulette wheel and retain the best individuals.

(5) Crossover operation. Assume x1 and x2 is parents, its children y1 and y2 after crossover is gained by the formula as bellow:

$$y_1 = \alpha x_1 + (1 - \alpha) x_2, \quad y_2 = \alpha x_2 + (1 - \alpha) x_1 \quad (3)$$

Here,  $\alpha$  is an parameter which changed with the evolution algebra.

(6) Variation. Take Gaussian approximate variance to improve the local search performance of GA on key search area. During the variation, using a random number of normal distribution which average is  $\bar{P}$  and variance is P2 instead of original gene.

(7) Fitness value of chromosomes group is calculated again.

(8) If it meets the stop search criteria, output the result. Otherwise, go to step (4).

### 3.3 Construction of RBF neural network model

Radial basis function network (abbreviations is RBFN) is a kind of neural network. The ability of function approximation, patten recognition and classification is better than BPNN [8-10], and is also used in lots of forecasting fields. Its structure is similar to BPNN. Input layer nodes only pass the signal to hidden layer. Activation function in hidden layer is radial basis functions. Generally, activation function in output layer is a simple linear function. In RBFN, input layer is mapped to a new space and achieves a linear combination. Adjustable parameters are the weights of linear combination and parameters which control the shape of the basis functions. The learning algorithm of RBFN has three parameters to solve which are the center of function, variance and the weights from hidden layer to out layer. In the phase of self-organizing learning, solve the center of basis function and variance in hidden layer. On instructor learning stages, solve the weights which are between hidden layer and output layer. Specific steps are as follow:

(1) Determination of basis function center C based on K-mean clustering

1) Network initialization. Training samples are selected randomly as the clustering center  $c_i$  ( $i = 1, 2, \dots, h$ ).

2) Training samples grouping. In accordance with the Euclidean distance between the training samples  $x_p$  and the center  $c_i$ , distribute the  $x_p$  to cluster set  $G_p$  ( $p=1, 2, \dots, P$ ) of input samples. Here, P is the total number of samples.

3) Re-adjustment of cluster centers. Calculate the average of training sample in crusting set  $G_p$  to gain the new clustering center  $c_i$ . If the new clustering center does not change any more, the  $c_i$  is the final center of the basic function. Otherwise, go to step 2).

(2) Variance  $\sigma_i$  Solution

The basic function of RBFN is Gaussian. So, the variance solution formula is as follows:

$$\sigma_i = \frac{c_{\max}}{\sqrt{2h}} \quad i=1,2,\dots,h \quad (4)$$

Here,  $c_{\max}$  is the maximum distance among selected centers.

(3) The weight calculation between hidden layer and output layer

The weights between hidden layer and output layer are calculated by least square method, the formula is as follow:

$$w = \exp\left(\frac{h}{c_{\max}^2} \|x_p - c_i\|^2\right) \quad p=1,2,\dots,P; \quad i=1,2,\dots,h \quad (5)$$

### 3.4 Integrated prediction model construction

The above models provide a new method for the prediction of agricultural production market price. However, one model more or less has shortcomings. If integrate multiple

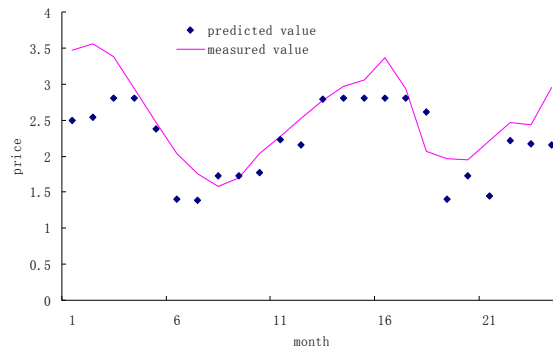
models, prediction process can be reflected from different sides to gain a accurate result. For the complexity of vegetables market price, It is difficult to describe the price variation rule using single model. So, the relationship between different forecasting model and final result is described by ANN to obtain the optimal forecasting result.

Firstly, predict the market prices of vegetables through separate model. Secondly, construct a three feed forward neural network model taking forecasting results as input value and actual result as output value. Finally, train the model using GA. When it is applied in real prediction, follow the same steps above. The training steps of integrated prediction model is the same with 3.2

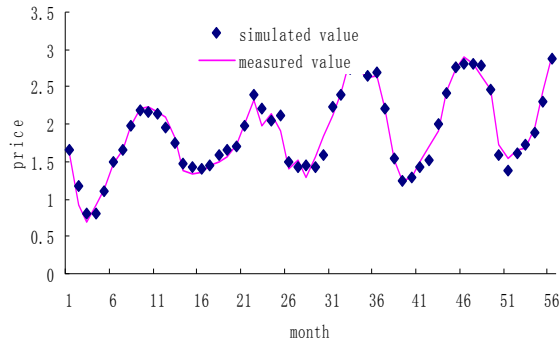
## 4 Results and discussion

### 4.1 Simulation and prediction analysis of BPNN

Take the former four month data of Lentinus edodes as input and the latter one month date as output. Code the program in Matlab. Construct BPNN. Use the data of 2003 – 2007 to train the model. The number of the hide neurons is 20. Circulation is 3000. Optimization target is 0.1. After training, the model simulates the price of 2003-2007, and predicts the price of 2008-2009. The results are as follow:



**Fig. 1.** Prediction result of BPNN

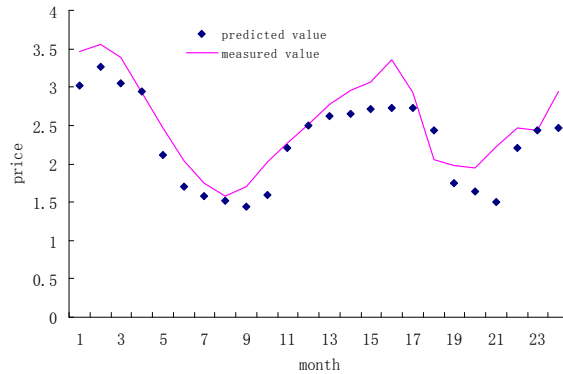


**Fig. 2.** Simulation result of BPNN

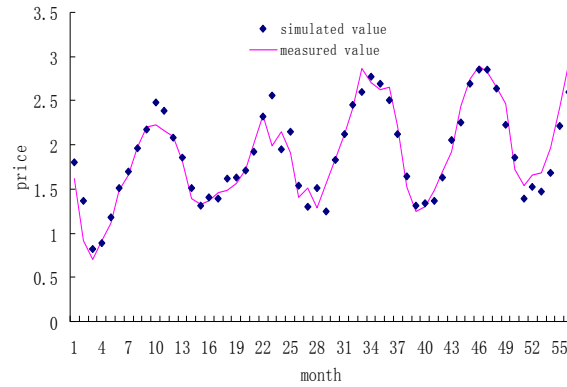
Calculate the mean absolute error (MAE) between measured value and simulated value, predicted value. In the figure above, the MAE of simulation is 0.046%, the MAE of prediction is 0.15%. Beside few point has greater error, on the whole, it is accurate on price trend prediction. It also indicates that BPNN has the reference value for the price prediction of Lentinus edodes.

#### 4.2 Simulation and prediction analysis of neural network Based on genetic algorithm

In this model, the chromosome group is 30, optimization goal is 0.5. Optimization algebra is 3,000. Its input value and output value are the same as BP neural network. The number of hidden layer is 15. Code the program in Matlab. Use the data of 2003 – 2007 to train model and the data of 2008-2009 to predict. The results are showed in fig.3 and fig.4:



**Fig. 3.** Prediction result of neural network based on GA

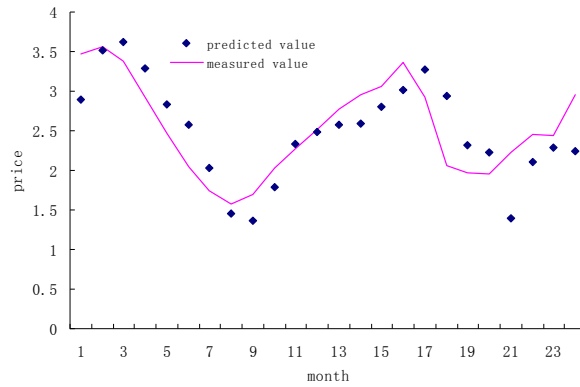


**Fig. 4.** Simulation result of neural network based on GA

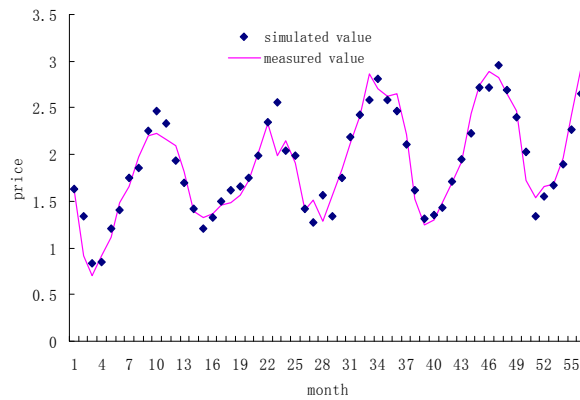
As can be seen from the figure above, the simulation results of neural network based on GA is worse than BP neural network. Its MAE is 0.075%.The prediction result of neural network based on GA is better than BPNN, the MAE is 0.114%.It shows that the simulation ability of neural network based on GA is worse than single ANN. However, its generalization ability is stronger than the neural network model, and with good results in forecasting.

#### 4.3 Simulation and prediction analysis of RBFN

The input value and output value of RBFN is similar to BPNN. The number of hidden layer neurons is 50. Optimization goal is 0.1.Code the program in Matlab. Use data of 2003 – 2007 to train model and the data of 2008-2009 to predict. The results are showed in fig.5 and fig.6:



**Fig. 5.** Prediction result of RBFN



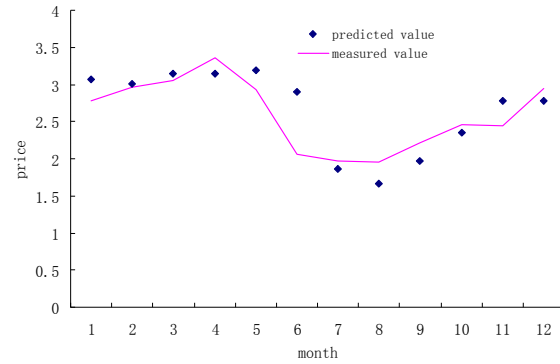
**Fig.6.** Simulation result of RBFN

In the figure above, the simulation results of RBFN is worse than BPNN and near by neural network based on GA. Its MAE is 0.071%. Its prediction result is closed to BPNN and worse than neural network based on GA. The MAE of prediction is 0.144%.

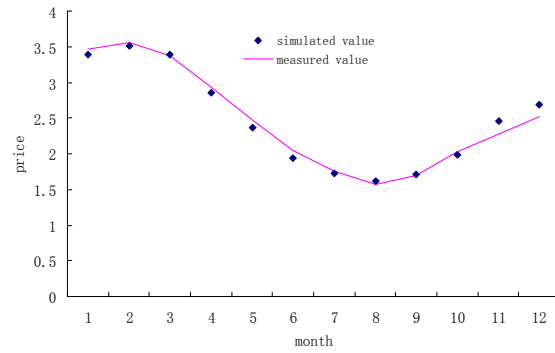
#### 4.4 Simulation and prediction analysis of integrated prediction model

Use the data of 2008 to train model and the data of 2009 to predict, the result is as below:





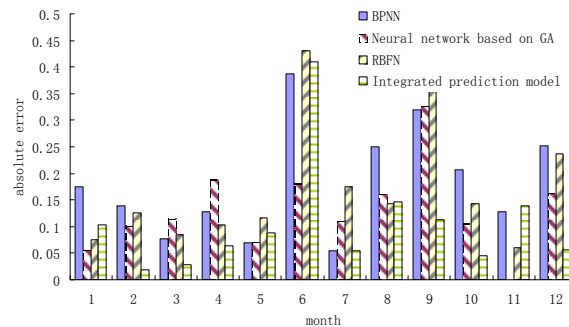
**Fig. 7.** Prediction result of integrated prediction model



**Fig. 8.** Simulation result of integrated prediction model

The result show that, using integrated prediction model, the MAE of simulation is 0.059%. It is better than neural network based on GA and RBFN, and worse than BPNN. The MAE of prediction is 0.106%. It is the best of all models.

Apply the four models to predict the market price of Lentinus edodes in 2009, the absolute error is showed in the fig.9:



**Fig. 9.** Absolute error of four models

The percentage of the number which predicting absolute error is greater than 20%, 15%, 10%, in the total predicting number, is displayed in table1:

**Table 1.** Comparison of predicting absolute error distribution of different model

model	>20%	>15%	>10%
BPNN	33.3%	50%	75%
Neural network based on GA	8.3%	41.6%	75%
RBFN	25%	33.3%	66.7%
Integrated prediction model	8.3%	8.3%	41.6%

As can be seen from the table, whenever in prediction, the results of integrated prediction model is equal or better than other three models. And then, the neural network based on GA is followed by. The performance of RBFN is worse than the neural network based on GA. BPNN is the worse of all.

## 5 Conclusions

To the vegetable price of Beijing wholesale Market, three models which are the BPNN, neural network based on GA and RBFN are established separately. Based on that, an integrated prediction model is constructed. Take Lentinus edodes price as experiment data, some conclusions is as follows:

(1) BPNN is good at simulation, but relatively poor at prediction. On the whole, it has accurate prediction on the trend of vegetable market price, and can offer a reference to prediction of Ledodes market price.

(2)The simulation ability of neural network based on GA is worse than BPNN. But its generalization ability is good, predicting accuracy is better than BPNN.

(3) The RBFN has a similar simulation capability with neural network based on GA. Its forecasting result is better than BPNN, and worse than neural network based on GA.

(4) The integrated prediction model utilizes the advantage which is provided by single predicting method, obtains the best prediction accuracy. It can improve the performance of agricultural market price prediction effectively.

The application based on the theory of ANN and GA can offer a good result for the agriculture production price predicting. In this paper, the price data itself is only considered to construct models. Further research is required for improving the accuracy of prediction. For example, more influential factors need to be considered, and the influence percentage of each factor needs more investigation.

## Acknowledgements

We are thankful that the study is supported by the Beijing Natural Science Foundation (9093019), and acknowledge the Beijing Municipal Bureau of Agriculture for providing the essential raw data.

## References

1. N. Amjady,: Day-ahead price forecasting of electricity markets by a new fuzzy neural network. IEEE Transaction on Power Systems. 887-896(2006).
2. T. Yamashita, K. Hirasawa and J. Hu: Multi-branch neural networks and its application to stock price prediction. Knowledge-based Intelligent Information and Engineering systems. Springer Berlin, Heidelberg. 2005
3. M.S. Wei, G.Y. Li: FNN-based Intelligent Insect Pest Forecast for Crops. Journal of Taiyuan University of Science and Technology 442-445(2007).
4. B. Najafi, M. Zibaei and MH Sheikhi, Forecasting price of some crop products in Pars province: application of artificial neural network. Journal of Science and Technology of Agriculture and Natural Resources. 501-512(2007).
5. C.S. Luo: Estimating Root density distribution of winter wheat under Water and salinity stress using the artificial neural network model based on Genetic algorithm, China Agricultural University 20-60(2002).
6. Q.W. Gu, G. Chen and L.L. Zhu: Short-term marginal price forecasting based on genetic algorithm and radial basis function neural network, Power System Technology 18-21, 25 7(2006).
7. C.S. Wu, C. Wu, L.S. Kang: Research on stock price forecasting methods by support vector machines based on genetic algorithms. 3rd International Conference on Innovation and Management. Wuhan PR China 29-30(2005).
8. L. Shao, X.D. Zhou: Industrial water demand forecast in Shanxi province based on RBFN. Yellow River 53-56(2010).
9. X.D. Qiang, Q. Xiao and H.Y. LUO: Research on prediction of RMB exchange rate based on improved RBF neural network, Computer Engineering and Applications 229-231(2010).

[10] J. Wang, L. Tian and H. Jiang: Forecasting of Short-term Power Load in RBF Network Based on Genetic Algorithm. *Electronic Technology* 15-1(2010).

