

Research Article

The Combination Forecasting Model of Grain Production Based on Stepwise Regression Method and RBF Neural Network

Lihua Yang and Baolin Li

School of Economics and Management, Hubei University of Automotive Technology,
Shiyan, 442002, China

Abstract: In order to improve the accuracy of grain production forecasting, this study proposed a new combination forecasting model, the model combined stepwise regression method with RBF neural network by assigning proper weights using inverse variance method. By comparing different criteria, the result indicates that the combination forecasting model is superior to other models. The performance of the models is measured using three types of error measurement, which are Mean Absolute Percentage Error (MAPE), Theil Inequality Coefficient (Theil IC) and Root Mean Squared Error (RMSE). The model with smallest value of MAPE, Theil IC and RMSE stands out to be the best model in predicting the grain production. Based on the MAPE, Theil IC and RMSE evaluation criteria, the combination model can reduce the forecasting error and has high prediction accuracy in grain production forecasting, making the decision more scientific and rational.

Keywords: Combination forecasting, grain production forecasting, RBF neural network, stepwise regression method

INTRODUCTION

Grain is an indispensable part of our life. With the rapid development of China's economy, China as a major production and consumption country of grain which faces with less and less arable land and increasing population, we must always pay more attention to grain production. Therefore, grain production forecasting is very significant, which can be used to make up the agricultural policy and adjust the structure of economy.

At present, many scholars have studied the model of grain production forecasting. For example, Bates and Granger (1969) proposed the combination forecasting model. Li and Ren (2009) discussed the crucial factors affecting grain production using the grey correlation and forecast grain production in Shanxi province using MOBP based on BP neural network of momentum method.

Dong *et al.* (2010) studied regression method based on SVM classification in the application of grain production forecasting. Liu *et al.* (2010) used multiple linear regression model based on forward selection variables method to predict grain production. Yao *et al.* (2010) compared four models in grain production forecasting, the results showed that BP neural network performed best. Chen (2011) predicted the unit grain yield of Hunan Province by using weighted Markov chain model, regression model and grey system model.

Niu (2012) presented an Adaptive Immune Genetic Algorithm optimization to optimize BP neural network's initial weights based on the concentration of the immune system regulation mechanism and Genetic Algorithm global optimization characteristics, for the BP neural network prediction of grain production fall into local optimum easily. Xiang and Zhang (2013) proposed a grain yield prediction model based on grey theory and markov model and the simulation results show that the proposed algorithm has better prediction accuracy.

However, there are still few studies that used combination forecasting model based on stepwise regression method and RBF neural network in the application of grain production forecasting and solved the problem using SPSS. This study will analyze this problem by the actual agricultural data.

SELECTION OF GRAIN PRODUCTION DATA

Grain production is a very complex biological and ecological process. The factors affecting grain production forecasting include advancement of science and technology, agricultural policies, natural factors and so on and there exist complex relationships between various factors.

As shown in Table 1, Grain production data which are selected from the China Statistical Yearbook are applied in the experimental analysis. We choose grain

Corresponding Author: Baolin Li, School of Economics and Management, Hubei University of Automotive Technology, Shiyan, 442002, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

Table 1: Grain production and affecting factors

Year	Actual grain production (10, 000 tons) Y	Sown area (10, 000 hm ²) ×1	Fertilizer dosage (10, 000 tons) ×2	Disaster area (10, 000 hm ²) ×3	Coverage area (10, 000 hm ²) ×4
1981	32502	11496	1334.9	1874.3	1.5
1982	35450	11346	1513.4	1598.5	11.8
1983	38728	11405	1659.8	1620.9	62.9
1984	40731	11288	1739.8	1560.7	133.4
1985	37911	10885	1775.8	2270.5	144.5
1986	39151	11093	1930.6	2365.6	179.6
1987	40298	11127	1999.7	2039.3	232.8
1988	39408	11012	2141.5	2394.5	229.5
1989	40755	11221	2357.1	2444.9	265
1990	44624	11347	2590.3	1781.9	331.1
1991	43529	11231	2805.1	2781.4	477.4
1992	44266	11056	2930.2	2585.9	568.9
1993	45649	11051	3150.1	2313.3	572.2
1994	44510	10854	3317.9	3138.3	624.4
1995	46662	11006	3593.7	2226.7	649.3
1996	50454	11255	3827.9	2123.3	744.9
1997	49417	11291	3980.7	3030.9	915
1998	51230	11379	4083.7	2518.1	1003.5
1999	50839	11316	4124.3	2673.1	1011.5
2000	46218	10846	4146.4	3437.4	1062.5
2001	45264	10608	4253.8	3179.3	1096.1
2002	45706	10389	4339.4	2731.9	1170.1
2003	43070	9941	4411.6	3251.6	1196.7
2004	46947	10161	4636.6	1629.7	1306.3

sown area ×1, fertilizer dosage ×2, disaster area ×3, coverage area ×4 as four factors affecting grain production forecasting (Dong *et al.*, 2010; Ding, 2007). The data from 1981 to 2000 in China are used to create the training sample sets. Grain production data from 2001 to 2004 in China are used to test the prediction ability of the proposed model.

COMBINATION FORECASTING MODEL IMPLEMENTATION

Stepwise regression method: Stepwise regression method is known as step by step selection method. It selects and removes variables repeatedly. Finally, the significant crucial variables are selected by the method. At each step, the independent variable not in the equation that has the smallest probability of F is entered, if that probability is sufficiently small. Variables already in the regression equation are removed if their probability of F becomes sufficiently large. The method terminates when no more variables are eligible for inclusion or removal.

Here, the actual grain production y is a dependent variable, sown area ×1, fertilizer dosage ×2, disaster area ×3 and coverage area x4 are independent variables. These variables are analyzed by stepwise regression method with the help of SPSS16.0 software. Variables can be entered or removed from the model depending on the significance of the F value. A variable is entered into the model if the significance level of its F value is less than the Entry value 0.05 and is removed if the significance level is greater than the removal value 0.10. The final result shows that ×2 and ×3 are crucial

factors highly affecting y. The multiple regression model formula is expressed as followings:

$$Y = 32822.587 + 6.129 \times 2 - 2.805 \times 3$$

RBF neural network: A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification and system control. RBF neural network is introduced to improve the performance of BP neural network (Modai *et al.*, 2002). Compared with BP neural network, RBF neural network employs RBF hidden units which make RBF neural network have more nonlinear modeling ability. RBF neural network is composed of an input layer, a hidden layer with a non-linear RBF activation function and an output layer, The input can be modeled as a vector of real numbers, The output of the network is then a scalar function of the input vector. In addition to the unnormalized architecture, RBF networks can be normalized, RBF network can solve over-fitting and local extremum of traditional neural network (Park, 2013).

Firstly, the experimental data should be normalized to get rid of the affection of scale. We usually can obtain a good training effect when values of input vector and output vector are normalized into a small scale. Here, we use Standardized method for rescaling dependent variable and covariates in order to prevent having bad generalization ability and that is subtract the mean and divide by the standard deviation, $(x - \text{mean})/s$.

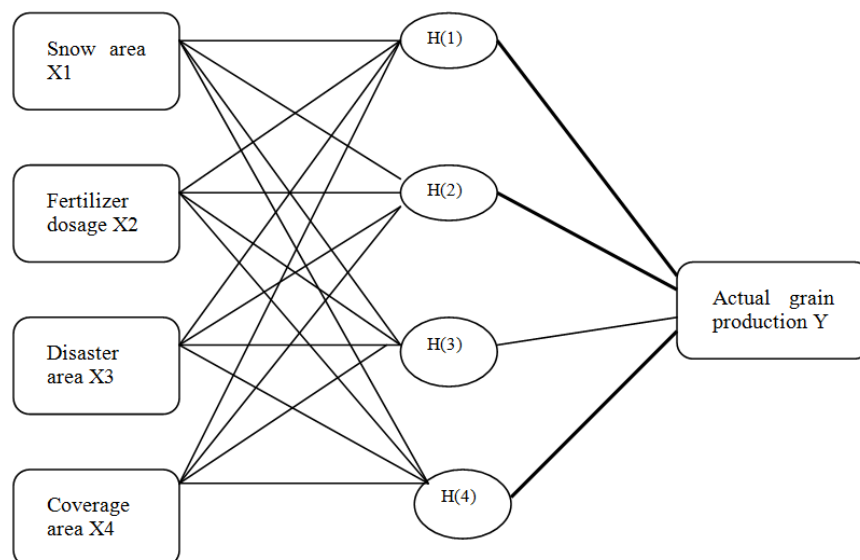


Fig. 1: The structure of RBF neural network

Table 2: Related parameter estimates in RBF

Predictor		Hidden layer				Output layer
Predictor		H(1)	H(2)	H(3)	H(4)	Actual grain production Y
Input layer	Sownarea×1	1.046	-0.560	-0.873	0.702	
	Fertilizerdosage×2	-1.009	-0.729	0.590	1.288	
	Disasterare×3	-1.259	-0.070	0.789	0.478	
	Coverage area×4	-1.003	-0.713	0.563	1.301	
Hidden unit width		0.470	0.458	0.835	0.510	
Hidden layer	H(1)					-0.922
	H(2)					-0.846
	H(3)					0.410
	H(4)					1.618

The structure of the RBF neural network is given in Fig. 1.

The number of units in the input layer is four, that is the number of covariates ×1, ×2, ×3 and ×4. RBF hidden layer has four units. The best number of hidden units is the one that yields the smallest error in the testing data. The activation function for hidden layer is softmax, so the activations of all hidden units are normalized to sum to 1. For the output layer, the activation function is the identity function; thus, the output units are simply weighted sums of the hidden units. The related parameter is shown as Table 2.

COMBINATION FORECASTING MODEL

Information is different to be used in each single forecasting model and its accuracy is also different. The information, each forecasting model contains, is fully used in combination forecasting model, through integrated them, unique uncertainties in single forecasting model are detracted, the overall uncertainty can also be reduced. Combination forecasting was

proposed by Bates and Granger (1969), they advocated the use of weighted average in the combination of forecasts. The basic principle of combination forecast model is entrust single forecasting model result with appropriate weight, thus combines a sole compound model. How to extract combination weight is the most essential question (Lu *et al.*, 2003).

The weights are calculated from the variance and covariance of different forecast errors in this study. The weighted combination is shown to be more accurate than any one of them. The basic principle of combination forecasting is to give the proper weight combination into a single composite model from the results of each single forecasting model. There are many different weight calculation methods, inverse variance method is selected:

$$W_i = D_i^{-1} / \sum_{i=1}^m D_i^{-1}, i = 1, 2, \dots, m$$

$$D_i = \sum_{j=1}^n (y_j - \bar{y}_j)^2$$

where,

m = The number of model

w_i = The weight of the ith model

D_i = The sum of square error of the ith-model

y_j = The actual grain production

\bar{y}_j = The predictive value

By calculating, the weight of the stepwise regression method is 0.036 and that of the RBF neural network is 0.964. The Combination forecasting model formula is expressed as followings:

$$Y = 0.036*y_1 + 0.964*y_2$$

where,

y₁ = The predictive value of stepwise regression method

y₂ = The predictive value of RBF neural network

Y = The predictive value of combination forecasting model

EXPERIMENTAL RESULTS

In the experiment, we respectively use stepwise regression method, RBF neural network model and combination forecasting model to establish single forecasting model. The actual grain production and the forecasting production from 2001 to 2004 are compared, The results are shown in Table 3.

Many measures of forecast accuracy have been proposed in the past, for example, using MAPE as forecast accuracy, Marianela use the root mean square percent error as forecast accuracy, Armstrong and Collopy recommended to use geometric mean relative absolute error, median relative absolute error and median absolute percentage error, median relative absolute error, symmetric mean absolute percentage error and symmetric median absolute percentage error were used, Fildes suggested using median absolute percentage error and geometric mean relative absolute error, Hyndman and Koehler proposed that the mean absolute scaled error become the standard measure for comparing forecast accuracy across multiple time series (Chen, 2010).

Model comparison methods generally include the absolute value and relative value comparison, The mean absolute error and root mean square error belong to the absolute value comparison, The theil inequality coefficient and mean absolute percentage error belong to the relative numerical comparison.

Here, we choose several criteria to evaluate the performance of the three above models. Three types of error measurement are used in this study, which are MAPE, Theil IC and RMSE (Ding, 2007). The MAPE is Percentage of the mean ratio of the error to the original data, which has been widely used as a performance measure in forecasting, the major reason

Table 3: Forecasting results of different models

Year	Actual grain production	Regression model	RBF model	Combination model
2001	45264	49976.1907	45261	45430.7469
2002	45706	51755.7901	45261	45494.8124
2003	43070	50740.5454	45261	45458.2636
2004	46947	56668.9999	45261	45671.6880

Table 4: Comparison of different models

Model	MAPE	Theil IC	RMSE
Regression model	15.5412	0.0746	7282.8437
RBF model	2.4146	0.0155	1400.0992
Combination model	2.273	0.015	1360.3886

for its popularity is that it is easy to interpret and understand. The range of Theil IC is from zero to one, if the fitting values and true values is smaller, then the differences is smaller. RMSE is Square root of the average of the summing square forecasting errors.

The related formulae are expressed as followings:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\bar{y}_i - y_i}{y_i} \right| \times 100$$

$$TheilIC = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^{-2} + \frac{1}{n} \sum_{i=1}^n y_i^2}}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2}$$

where,

\bar{y}_i = Forecast grain production

y_i = Actual grain production

i = The ith year

The comparison results are shown in Table 4.

From the Table 4, we can see that the combination model based on stepwise regression and RBF neural network has best accuracy, because the responding values of MAPE, Theil IC and RMSE are the best among the three models. The RBF neural network has the better accuracy than stepwise regression.

In the process of combination, each single model's advantages strengthened and disadvantages weaken. Therefore, combination forecasting model has higher accuracy and reliability by integrating useful information of each single model losing in the each single forecasting model.

CONCLUSION

According to the actual grain production data from 1981 to 2000 in China, through analysis of four factors affecting total grain production with the method of stepwise regression analysis, the result indicates that two major factors are fertilizer dosage and disaster area.

At present, China's population has been controlled effectively. Increasing total grain production is an important way to ensure food security. The fertilizer dosage and disaster area are two significant factors which affect actual grain production y . The two significant factors should be the core point of the agriculture work. At the same time the government must increase investment in agricultural science and technology. The government should actively develop and promote advanced agricultural technologies so as to decrease the disaster area.

The paper respectively makes use of stepwise regression method, RBF neural network model and combination forecasting model to establish forecasting model, the RBF neural network model which is established by sown area, fertilizer dosage, disaster area, coverage area and actual grain production has the precise prediction result and allocate the proper weight by using inverse variance method. Based on combination forecasting theory, the novel combination forecasting model of the grain production has been established and we apply this kind of model to forecast the grain production during the following 4 years.

The numerical example proved that the combination forecasting model of grain production based on stepwise regression method and RBF neural network can overcome the single model in previous decision makings, it can reduce the forecasting error and has high prediction accuracy, providing the decision support for grain production forecasting.

ACKNOWLEDGMENT

The study has been supported by project of the Department of Education, Hubei Province China (NO. B20092306) and by project of Hubei University of Automotive Technology (No. 2012XQ02).

REFERENCES

- Bates, J.M. and C.W.J. Granger, 1969. The combination of forecasts. *Oper. Res. Quart.*, 20: 451-468.
- Chen, H.J., 2010. A new measure of forecast accuracy. *Proceeding of the 2nd IEEE International Conference on Information and Financial Engineering (ICIFE, 2010)*, 1: 710-712.
- Chen, J.B., 2011. Prediction of unit grain yield based on grain safety level in Hunan province. *Syst. Eng.*, 5: 113-117.
- Ding, C.F., 2007. Application of combination forecasting method in grain production forecast. *Res. Agric. Mod.*, 28: 101-103.
- Dong, Y., W. Cheng, Y.P. Zhang and S. Zhao, 2010. Regression method based on SVM classification and its application in production forecast. *J. Comput. Appl.*, 30: 2310-2313.
- Li, H. and Z.Y. Ren, 2009. The Grey Correlation Analysis of factors affecting grain production and the forecast for grain production in Shanxi Province. *Territory Nat. Resour. Study*, 4: 56-58.
- Liu, D., X.F. Bai and J. Meng, 2010. Multiple linear regression forecasting model of total food yield in China based on forward selection variables method. *J. Northeast Agric. Univ.*, 41: 124-128.
- Lu, Q., P.L. Gu and S.M. Qiu, 2003. Construction and application of combination forecasting model in Chinese energy consumption system. *Syst. Eng. Theory Pract.*, 3: 24-28.
- Modai, I., M. Ritsner, R. Kurs, S. Mendel and A. Ponizovsky, 2002. Validation of the computerized suicide risk scale: A back propagation neural network instrument. *Eur. Psychiat.*, 17: 75-81.
- Niu, Z.X., 2012. Prediction of grain yield using AIGA-BP neural network. *Comput. Eng. Appl.*, 48: 235-237.
- Park, J., 2013. Universal approximation using radial-basis-function networks. *Neural Comput.*, 2: 246-257.
- Xiang, C.S. and L.F. Zhang, 2013. Grain yield prediction model based on grey theory and markvo. *Comput. Sci.*, 40: 245-248.
- Yao, Z.F., X.T. Liu, F. Yang and M.H. Yan, 2010. Comparison of several methods in grain production prediction. *Agr. Res. Arid Area.*, 28: 264-268.