A Prediction Model of Peasants' Income in China Based on BP Neural Network

GUO Qing-chun[1*], HE Zhen-fang[2], LI Li[3], KONG Ling-jun[1], ZHANG Xiao-yong[1], KOU Li-qun[1]

1. Shaanxi Radio &TV University, Xi'an 710068, China; 2. Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lan'zhou 730000, China; 3. Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710075, China

Abstract According to the related data affecting the peasants' income in China in the years 1978 –2008, a total of 13 indices are selected, such as agricultural population, output value of primary industry, and rural employees. Based on the standardized method and BP neural network method, the peasants' income and the artificial neural network model are established and analyzed. Results show that the simulation value agrees well with the real value; the neural network model with improved BP algorithm has high prediction accuracy, rapid convergence rate and good generalization ability. Finally, suggestions are put forward to increase the peasants' income, such as promoting the process of urbanization, developing small and medium-sized enterprises in rural areas, encouraging intensive operation, and strengthening the rural infrastructure and agricultural science and technology input.

Key words BP Neural Network, Peasants' income, Forecast, China

There are many research achievements about peasants' income. Xiong Zhanglin *et al*. take 31 provinces, autonomous regions, and municipalities as the samples, select disaster area, cultivated land area, fertilizer input and other 13 factors affecting the income of peasants, and use factor analysis method to analyze the influencing factors of peasants' income. Result shows that the common factors affecting the peasants' income are the expanded reproduction factor, agricultural production factor, information utilization factor, and reduction factor. The peasants' education degree and the peasants' ability of grasping the information have relatively great impact on peasants, and can effectively promote the income increase of peasants[1]. Zheng Sufang *et al*. select indices affecting the peasants' income in China, such as agricultural output, price of agricultural products, and per capita sowing area, use econometric model to find out that the major factors improving the peasants' income are the non-agricultural income of peasants and the price of agricultural products, based on the comprehensive analysis on various factors from the quantitative point of view[2]. Shen Juhong selects the related data affecting the peasants' income in Ningxia in the years 1990 –2005, establishes the principal component regression model about peasants' income, uses the data in the year 2005 to forecast, and obtains that the forecasting value is very close to the real value[3]. Wang Yana *et al*. use the stepwise regression method to analyze the factors affecting the peasants' income. Analysis result shows that the transfer of rural surplus labor force and the education level of peasants have the most significant impact on the per capita net income of peasants[4]. Previous researchers mainly use the linear mode, but the factors affecting the peasants' income are various; the income growth of peasants is complex with the characteristics of nonlinearity. Artificial Neural Network has strong self-learning ability, is good at generalization, association, analogy and reasoning, and can analyze macroscopic statistical laws from a large number of statistics. With strong fault tolerance and robustness and great parallel computing ability, Artificial Neural Network becomes an effective means for non-traditional modeling and is widely applied[5-8]. Therefore, based on the nonlinear characteristics of BP neural network, factors affecting the peasants' income are analyzed and suggestions to increase the peasants' income are put forward. Among them, the forecasting method of per capita net income in rural China based on the neural network with improved BP algorithm has made full use of the characteristics of high regression accuracy, good convergence and rapid arithmetic speed, as well as the strong nonlinear mapping capability of neural network. It overcomes the shortcomings of BP algorithm, improves the speed and accuracy of forecasts, and makes up for the shortages of biased estimation in the forecast of peasants' income by linear model. The research results can be used as references for the decision-making of the relevant government departments and are of great theoretical and practical values.

1 Index selection, data source and research method

1.1 Index selection According to the previous research results, factors affecting the peasants' income (independent variables) in rural China are agricultural employees X_1 ($\times 10^4$), output value of primary industry X_2 ($\times 10^8$ yuan), employed persons in rural areas X_3 ($\times 10^4$), fixed asset investment in rural areas X_4 ($\times 10^8$ yuan), output value of farming, forestry,

Received: March 5, 2011 Accepted: March 25, 2011
Supported by the National Natural Science Foundation of China (0211003026/11220300).
* Corresponding author. E-mail: gqc@ieecas.cn

animal husbandry, and fishery X_5 ($\times 10^8$ yuan), agricultural machinery power X_6 ($\times 10^4$ kW), effective irrigation area X_7 ($\times 10^3$ hm^2), fertilizer input X_8 ($\times 10^4$ t), investment in hydropower construction in rural areas X_9 ($\times 10^4$ yuan), original value of productive fixed assets X_{10} (yuan/household), total sowing area of crops X_{11} ($\times 10^3$ hm^2), grain output X_{12} ($\times 10^4$ t), and disaster area X_{13} ($\times 10^3$ hm^2). The dependent variable is the per capita net income of peasant households Y (yuan).

1.2 Data source The research data are from the State Statistics Bureau.

1.3 Research method

1.3.1 Standardized method. Due to the different physical meanings of influencing factors, the index selected has no unified dimension and dimensionless unit, so that the dimensions of data are different. When inputting variables, the right conclusion can hardly be obtained. In order to make the index more comparable, dimensional treatment on data is needed before the application of neural network. Generally speaking, due to the great data span colleted, the input data should be within [−1, 1] during the practical application of neural network, in order to effectively avoid the saturation region of Sigmoid function and to enhance the training speed.

1.3.2 BP neural network method

1.3.2.1 The fundamental principle of BP neural network. The neural network model of Error Back Propagation (BP) algorithm is a multilayer feedforward network with one-way communication. Except the input layer, the network has one or more hidden layers and there is no coupling in the same level nodes. The input signal in turn passes through the nodes at hidden layer to the nodes at output layer. Node output at each layer only affects the output at the next layer. When given the network an input mode, it is transferred from the node at input layer to the node at hidden layer, and then to the node at output layer. Finally, after the treatment of node at output layer, an output mode is produced. If the expected output is not obtained at the output layer, we can turn to the reverse propagation process and the error signal is turned back through the original connection path. Based on modifying the weights and thresholds of neurons at different layers, calculation is carried out in turn to the output layer, and then the forward propagation process is carried out. Repeated use of the two processes can make error achieve the requirements of people; and then the learning process is ended. During the practical application, a set of training samples are prepared and each of them includes both input sample and ideal output samples. The training is ended when all the errors between network output and ideal output are within the expectation of people. Otherwise, the reverse propagation process of error is used to modify the weights and thresholds, and to make the network output consistent with the ideal output. Repetition training is needed until the overall error of sample reaches the accuracy that people expected; and the weights and thresholds after adjustment are kept for calculation.

1.3.2.2 Selection of input layer, output layer, and the nodes of hidden layer. A total of 13 indices affecting the per capita net income of peasant households are used as the input vectors of BP neural network. The per capita net income of peasant households in rural China is taken as the output vector. Therefore, there are 13 nodes at input layer and 1 node at output layer. Theory has proved that there is only one network at hidden layer that can reach the approximation of any rational function with an arbitrary order of accuracy, so that one hidden layer is selected during the network construction of the prediction model of per capita net income of peasant households in rural China. During the prediction of per capita net income of peasant households, BP network with 1 – 30 nodes at hidden layer is constructed. Based on pilot calculation, it is obtained that the idea hidden units are 9 according to the test results reported.

1.3.2.3 Determination of training function. Due to its disadvantages, BP neural network should be improved. In this research, trainlm (Levenberg-Marquardt optimization method), traingdm (momentum gradient descent back propagation algorithm), trainbr (Levenberg-Marquardt optimization method and Bayesian regularization method) are compared and analyzed. And then, we select the traingbr training method, which has relatively high fitting accuracy, and rapid convergence rate. Training performance function of modified neural network is used to improve the generalization ability and to avoid the overtraining. Modified algorithm of BP neural network is used to analyze the data of per capita net income of peasant households in rural China; and the prediction model based on BP neural network is established.

2 Result and analysis

2.1 Simulation results Data in the years 1978 – 2003 are used as the training samples; and data in the years 2004 – 2008 are used as the test samples. Simulation is carried out by the improved BP neural network. It should be noticed that after the training of pretreated data, inverse transform of the results of network output should be carried out in order to obtain the real value. Anti-preprocessing algorithm is $x = (y + 1) \times [\max(x) - \min(x)]/2 + \min(x)$. Table 1 reports the comparison between the real value and the simulation value. Y is the real value (yuan), Y_1 is the simulation value (yuan), R is the residual error (yuan), and RE is the relative error (%). In order to analyze the impacts of influencing factors on the per capita net income of peasant households, network training is used to obtain the weights between input layer and hidden layer, as well as the weights between output layer and hidden layer. According to the calculation methods by Gao Renxiang *et al.* [9], Table 2 reports the impact proportions of factors on the per capita net income of peasant households in rural China.

2.2 Simulation analysis Specific analysis includes the following three aspects:

(1) Table 1 shows that the average relative error between the real value and the simulation value in training sample is 1.0%, and the residual range is between −2.7% and 3.8%. The average relative error between the real value and the simulation value in test sample is 0.4% and the the residual range is between 10.8% and 14.6 % and the range of relative error is

between 0.2% and 0.5%. Thus, both the training sample and test sample have relatively small relative errors.

Table 1 Comparison between the real value and the simulation value

Year	Y	Y_1	R	RE
1978	133.6	132.5	1.1	0.823 4
1980	191.3	190.2	1.1	0.575 0
1985	397.6	398.3	−0.7	−0.176 0
1990	686.3	684.9	1.4	0.204 0
1991	708.6	701.7	6.9	0.973 8
1992	784.0	775.3	8.7	1.109 7
1993	921.6	912.8	8.8	0.954 9
1994	1 221.0	1 174.0	47.0	3.849 3
1995	1 577.7	1 565.8	11.9	0.754 3
1996	1 926.1	1 914.5	11.6	0.602 3
1997	2 090.1	2 013.6	76.5	3.660 1
1998	2 162.0	2 153.0	9.0	0.416 3
1999	2 210.3	2 120.4	89.9	4.067 3
2000	2 253.4	2 234.7	18.7	0.829 9
2001	2 366.4	2 351.2	15.2	0.642 3
2002	2 475.6	2 463.1	12.5	0.504 9
2003	2 622.2	2 692.6	−70.4	−2.685 0
2004	2 936.4	2 921.8	14.6	0.497 2
2005	3 254.9	3 243.5	11.4	0.350 2
2006	3 587.0	3 572.4	14.6	0.407 0
2007	4 140.4	4 127.5	12.9	0.311 6
2008	4 760.6	4 749.8	10.8	0.227 3

Table 2 Proportion of input variable to the output value in neural network

Y	Proportion	Y	Proportion
X_1	0.2	X_8	1.9
X_2	0.6	X_9	1.2
X_3	41.2	X_{10}	0.9
X_4	25.4	X_{11}	24.6
X_5	0.5	X_{12}	0.4
X_6	1.6	X_{13}	0.1
X_7	1.7		

(2) Table 2 shows that the proportions of employed persons in rural areas, the fixed asset investment in rural areas, and the overall sowing area of crops have the greatest proportion, which are 41.2%, 25.4% and 24.6%, respectively. Other influencing factors of peasants' income have relatively small proportions. Among them, the agricultural employees, the output value of primary industry, the grain output, and the disaster area have the proportions of 0.2%, 0.6%, 0.4%, and 0.1%, respectively. Proportion of employed persons in rural areas is the greatest, because the income source of peasants has changed from the household operating income in the past to the payment of labour. The fixed asset investment in rural areas has relatively great impact on the income of peasants. This is because peasants lacked the inner dynamics of using advanced agricultural technology in the past, the capital and technology input were replaced by the labor input, and the extensive management was carried out by traditional techniques and experiences. In recent years, development and implementation of advanced and useful agricultural technologies have greatly enhanced the efficiency of agricultural production. The overall sowing area of crops has relatively great impact on peasants' income, because the sowing area of crops has an important influence on the agricultural sustainable development, grain security, and ecological environment of China.

3 Policy suggestion

According to the result of neural network training, it can be seen that the income of peasants is mainly affected by the employed persons in rural areas, the fixed asset investment in rural areas, and the overall sowing area of crops. Since there are also problems such as insufficient selection of indices, the following suggestions are put forward in order to increase the income of peasants.

3.1 Promoting the process of urbanization Promoting the process of urbanization can properly arrange the rural surplus labor force, alleviate the conflict between many rural populations and little arable land, create more job opportunities for the rural surplus labor forces, and increase the income of peasants. Urbanization can stimulate the investment and consumption needs of rural residents, solve the problems of low labor capacity and insufficient demand to a certain extent, and speed up the economic development. Moreover, urbanization of rural areas can promote the structure upgrading and cluster development of rural industries, and create conditions for the transform of agricultural production mode and the intensive development of agriculture.

3.2 Developing small and medium-sized enterprises in rural areas Developing small and medium-sized enterprises can combine the rural industrialization and urban industrialization together. At present, peasants work in small and medium-sized enterprises and their stable income has become an important economic source of rural households. Besides, small and medium-sized enterprises in rural areas is rooted in the rural areas, can serve the peasants, is closely related to the agriculture, peasants and rural areas, and is the most effective, stable and convenient channel for the transfer of rural surplus labor force.

3.3 Encouraging intensive operation Encouraging intensive operation can circulate the land into the large households with operating skills and business experience. On the one hand, it can improve the agricultural productivity, enhance the economies of scale, and promote the development of agricultural industrialization. On the other hand, it can free some of the labor forces from the rural areas, and let them work in secondary industry or tertiary industry.

3.4 Strengthening the rural infrastructure and agricultural science and technology input Firstly, we should enlarge the total investment, improve the road, water conservancy and other infrastructures in rural areas, and enhance the condition of agricultural production. Secondly, we should change the investment structure, increase the input in agricultural technology and agricultural industrialization, give more finance support for the

(To page 94)

tion, so they are more open-minded and they fight more apparently against the prevalence and harm of the man-oriented culture. At the macro level, the demand of the equal right of man and woman may get the social respect usually, but at the career level, the new generation migrant workers may confront with the career plights because of the suppress of man-oriented social structure. According to the relevant report, part of new generation migrant workers feel they do not have enough space in the job-market dominated by man, so they often down to the pornography industry or some other informal or illegal industry. For these career lets them feel that they are equal to man and they can compete with man equally in certain degree at least. It can be concluded that, the man-oriented social atmosphere has increased the uncertainty, unsafe and low security factors of new generation migrant workers. And thus the female workers have to bear more individual risks when realizing the rural-urban migration[8].

Except for the career plight, some social problems such as courtship problems, marriage problems, pregnancy plights and some other problems of new generation migrant workers caused by the gender difference will directly relate to whether we can construct the moral framework which is in line with humanity and whether we can lift the social culture level to the modern height. From this perspective, the problems concerning female workers are of great importance to the new generation migrant workers.

4 Conclusions

The problem of migrant workers, characterized by the Chinese feature, is the problem formed in a certain history environment since Chinese reform and opening up. There is no experience in the foreign countries to learn from directly, so the government should not only take the spot survey as the basis and the creative thoughts as the research direction, but also rely on the noble humanity to guide the stipulation of relevant social policies. The research analyses the social problems concerning " migrant workers" from the three aspects of " social discrimination, class mobility and gender direction". The research thinks that the migrant workers should appeal to the Chinese social environment from at least the three follow aspects: firstly, it is not only the system problem but also the culture problem, for the migrant workers who account for 10% of the total population of the whole country have affected by the social discrimination. It shows that there is still room for improving the whole national quality. Secondly, the plights confronted by the new generation migrant workers related to the complicated social problems concerning the urban and rural discrepancy, registration system, educational system, widening divide between the rich and the poor, which put the social injustice to an endangering position. Thirdly, understanding and smooth of the social plight of new generation migrant workers are the objective requirement of Chinese fine culture and it is testing the conscience and wisdom of Chinese society.

References

[1] HE HH, PI XP. Dialectical thinking on the concepts of migrant worker [J]. Truth Seeking, 2006(5): 56 -58. (in Chinese).

[2] ZHAGN YJ, JIANG ZH. Definition of "migrant worker" and its characteristics [J]. Jiangnan Forum, 2007(8): 16. (in Chinese).

[3] LIU ZQ. Protection of human rights and national obligation—from the perspective of the new generation of migrant workers [J]. Southeast Academic Research, 2010(2): 13 -20. (in Chinese).

[4] HE SW. Poverty is being inherited [J]. Tribune of Villages and Townships, 2006(8): 20. (in Chinese).

[5] ZHU SB. Marginal psychological is inherited [J]. Monthly Essays (original version), 2008(1): 25. (in Chinese).

[6] CAI H, LIU LP, WAN XD, *et al*. Migrant workers in the urbanization process—a study from the Pearl River Delta Area [M]. Beijing: Social Sciences Academic Press, 2009: 387. (in Chinese).

[7] The Research Team of State Council Research Office. Research report on Chinese migrant workers [R]. Beijing: China Yanshi Press, 2006: 86. (in Chinese).

[8] QIAN XF. Personal risk cost on the farmer -workers in rural to urban migration—based on the questionnaires of 1012 famer - workers [J]. Reformation & Strategy, 2008(2): 82 -85. (in Chinese).

(From page 90)

circulation after production and the information system construction before production, enhance the agricultural science and technology content, fundamentally enhance the competitiveness of agricultural markets, and raise funds to the agricultural technology extension and the agricultural researches in order to enhance the development potential.

References

[1] XIONG ZL, LIU Y. Analysis of the influencing factors of farmer´s income[J]. Journal of Anhui Agricultural Sciences, 2010, 38(20): 11030 -11032. (in Chinese).

[2] ZHEN SF, ZHANG YH. Analysis and countermeasures on influencing factors of farmer´s income in China[J]. Guangdong Agricultural Sciences, 2010(10): 223 -225. (in Chinese).

[3] SHEN JH. A principle component regression model to predict farmers´income[J]. Journal of Anhui Agricultural Sciences, 2008(24): 10285 -10287. (in Chinese).

[4] WANG YN, ZHA QF, SONG LP. Empirical analysis on the foctors of effecting farmer´s income[J]. Journal of Anhui Agricultural Sciences, 2007(1): 262 -263. (in Chinese).

[5] WANG RM, FU ZT, HE YY. Prediction model of dissolved oxygen fuzzy system in aquaculture pond based on neural network[J]. Journal of Anhui Agricultural Sciences, 2010(33): 18868 -18870, 18873. (in Chinese).

[6] ZHANG HM, SUN SQ. Quality evaluation model of atmospheric environment based on BP neural network[J]. Journal of Anhui Agricultural Sciences, 2010(31): 17656 -17657, 17664. (in Chinese).

[7] LV YJ, BAO Y, YANG DG. Application of artificial neural network in the climatic suitability evaluation of maize[J]. Journal of Anhui Agricultural Sciences, 2009(35): 17793 -17794. (in Chinese).

[8] REN XH, WANG YB, ZUI JX. Trapping number forecast of thrips alliorum by colorful plate based on BP neural network model[J]. Journal of Anhui Agricultural Sciences, 2009(31): 15294 -15296. (in Chinese).

[9] GAO RX, ZHANG SY, LIU B. Variable selection approach based on neural network analysis[J]. Journal of Systems Engineering, 1998, 13(2):32 -37. (in Chinese).