

Prediction of Visitors Quantity Based on A Combined Method

Yewang Zhou

School of Business, Huanggang Normal University, Huanggang, 438000, China

Abstract: Visitors quantity is regarded as the important economic indicator of weighing a country or regional tourism economic development, how to correctly and effectively predict regional visitors quantity according to relevant influencing factors is essential to the development of tourism economy. This paper firstly introduce several common prediction method principle, then according to the characteristics of visitors quantity, build combined prediction model based on neural network, Use fitting value got in front of three kinds of prediction method as input samples, use the corresponding to the actual value as the output samples, training the network learning, Amplify neural network to identify the weight of every prediction method, reduce subjective factors in the prediction process. The result shows that the method is feasible and effective, the results is accurate.

Keywords: Visitors quantity, prediction, combined method

I. INTRODUCTION

Prediction is the foundation and basis of decision. Using an appropriate prediction method to get tourism economy development rule and provide scientific basis for decision-making in the course of tourism economic development. Avoid economic and environmental loss caused by decision-making errors, it is necessary to realize the sustainable development of tourism, especial in regional tourism planning. Tourism market prediction is scientific or not, related to the tourism infrastructure, service facilities and ancillary facilities such as fixed asset investment and civil engineering is adequate, reception capacity is required, which is related to the whole tourism economy operation process. Therefore, the scientific tourism prediction play a decisive role in the begin development of tourism economy.

The current domestic visitors quantity prediction method can be divided into the following three categories: one is the extrapolation method, namely using past data to predict the future state, such as time series analysis; the second method is the law of cause and effect, according to the historical data, to find the relation between the variables to be forecasted and related variables, so as to predict the future state, as the regression analysis method; the third method is judgment analysis method, which mainly depends on the expert experience and comprehensive analysis ability to predict future states of [3].

Because of the diversity and complexity of tourism system influencing factors, although the current prediction methods have their advantages, but the data fit is not high, the prediction results have deviation from the actual value. Such as time sequence prediction method, using historical data of visitors quantity to predict the time trend, which does not take relative influencing factors into account; regression analysis method establish regression model based on some influencing factors that influence the amount of visitors quantity, but it only consider several influencing factors, Simulation ability to practical problems is not accurate. With the exception of the above methods, domestic scholars are still using some new artificial intelligence method to predict the visitors quantity.

As everyone knows, tourism science is a young discipline with only 20 years history in our country; it is still a difficult area to predict the tourism market. Compared with other industries, tourism industry has significantly volatility, and is easily affected by economic environment, political environment, unexpected incidents. Tourism is subject to the tourists' personal preferences, disposable income, paid time, enterprise marketing investment, product innovation and other factors in microcosm, it is difficult to strictly distinguish between endogenous and exogenous variables of tourism economic development. In addition, due to the development lag in China statistical system, many relevant statistical data about the development of tourism economy cannot be obtained. Therefore, research on the tourism market belongs to the "partial

information known, partly information unknown" "small sample, poor information, uncertain" system research. Based on this, this paper try to use multiple method mixed grey neural network and times series analysis to predict visitors quantity, then take Huanggang city as a example to demonstrate the feasibility of method.

II. METHODS

A. Artificial neural network

Artificial neural network (Artificial Neural Network ANN) is a large number of simple processing units connected neurons through extensive artificial network, the network can be used to simulate the brain nervous system function and structure. It can be known from data automatically induced rules, obtain the inherent laws of these data, and has strong nonlinear mapping ability, especially in the curve fitting with high accuracy.

At present, in the applications of the artificial neural network, the vast majority of neural network models are used in the error back propagation learning algorithm of feed forward neural network. BP neural network algorithm is as follows: The guiding ideology of BP neural network is: correct network weights (w_{ij}, T_{li}), and threshold (θ), make difference function (E) drop along the negative gradient direction, Three layer of BP network nodes is represented as, input node: x_j , implicit node y_i , the output node o_l , the weight of the network between the output node and hidden nodes is w_{ij} , the weight between hidden nodes and the output node is T_{li} , the desired output the of output node is t_l , The basic calculation formula of BP algorithm is as follows:

The hidden node output

$$y_i = f \left\{ \sum_j w_{ij} x_j - \theta_i \right\} = f(net_i) \quad (1)$$

The output of output node

$$o_l = f \left\{ \sum_i T_{li} y_i - \theta_l \right\} = f(net_l) \quad (2)$$

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

The relationship between

$$f'(x) = f(x)(1 - f(x))$$

$$\text{Then } f'(net_k) = f(net_k)(1 - f(net_k))$$

Output node

$$f'(net_l) = o_l(1 - o_l)$$

Hidden node

$$f'(net_i) = y_i(1 - y_i)$$

$$\text{Error control } E = \sum_{k=1}^p e_k < \varepsilon$$

$$\text{Error formula } \partial_l = (t_l - o_l)o_l(1 - o_l)$$

$$\text{Weighting formula } T_{li}(k+1) = T_{li}(k) + \eta \partial_l y_i$$

$$\text{Threshold formula } \theta_l(k+1) = \theta_l(k) + \eta \partial_l$$

B. Time series regression analysis prediction method

When predicting visitors quantity in the use of time series regression analysis prediction method, firstly we should analyze the relation of visitors quantity between independent and dependent variables correlation, and then established the regression equation between variables, and use the regression equation as predictive model, predict the dependent variable relationship in prediction of the quantitative change according to the independent variables. Regression analysis prediction method has many types, according to the correlation between the number of independent variables of different classification, can be divided into unary regression analysis and multiple regression analysis prediction method, In a regression analysis prediction method of arguments is only one, while in the multiple regression analysis prediction method, has more than two independent variables. According to different relationship between the independent variables and dependent variables also can be divided into linear regression and nonlinear regression. Considering internal connection and law about the data of visitors quantity, and the law will affect the later development, this paper takes the time series regression analysis method, which is the application of past visitors quantity to speculate about the future development trend.

C. Grey prediction method

When predicting visitors quantity in the use of grey prediction method, we use grey model (GM (1, 1) to solve sequence on the gray system. As a result of application of gray model data generating means, weakening the randomness of the disordered system, and make the original sequence showing a rule, Make no obvious rule becomes more obvious, and the model can identify on residual error. Using this method, even in less visitors quantity data, and data

is in arbitrary random distribution situation, the model also can get a higher precision of prediction.

If given original data series $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, These data are irregular, random, have obvious swing, if accumulate the original data and generate new data column, $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, in the above equation, $x^{(1)}(i) = \sum_{k=1}^i x^{(0)}(k), i=1, 2, \dots, n$. The new generation of data is a monotonic growth curve, increased the original data regularity, and weaken the fluctuation.

Grey system model is applied directly to the time series into a differential equation, namely, the abbreviation of grey dynamic model is GM, GM (1,1) contains a variable differential equation of first order dynamic model, GM (1,1) model of first order differential equation is:

$$\frac{dX^{(1)}(t)}{dt} + aX^{(1)}(t) = u \quad (3)$$

Use the least square method we can get

$$\hat{a} = \begin{vmatrix} a \\ u \end{vmatrix} = (B^T B)^{-1} B^T X_n \quad (4)$$

In the equation,

$$B = \begin{vmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(X^{(1)}(n-1) + X^{(1)}(n)) & 1 \end{vmatrix} \quad (5)$$

$$X_n = \begin{vmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \dots \\ X^{(0)}(n) \end{vmatrix} \quad (6)$$

$$X^{(1)}(K+1) = (X^{(1)}(0) - u/a)e^{-ak} + u/a \quad (7)$$

D. Combined prediction theory

Combined prediction method is to build a model, which is a prediction method use two or more than two different prediction methods yielded different predictive value by appropriate weighted average, finally take the weighted average value as the final prediction result. For a prediction problems, in a period of time the actual value, x_t , Combined prediction value y_t ($t=1, 2, \dots, n$), the combined prediction model can be expressed as:

$$\begin{cases} y_t = \sum_{i=1}^m a_i u_{ti} \\ \sum_{i=1}^m a_i = 1 \end{cases} \quad (8)$$

Obviously, if want to make full of the advantages of each single prediction method, we should well identify weighted coefficient of various prediction methods. Currently there are many methods to determine weighted coefficient of the various single prediction method such as mathematical programming methods, neural network method, and the paper select the neural network method to determine the weighted coefficient.

III. RESULTS

E. Neural network prediction

When predicting in the use of neural network prediction, through repeated comparison and research combined with expert advice, ultimately determine the input layer of the neural network is 3, the output layer is 1, and the training samples divide into 10 groups, in 2000 ~ 2002 data to predict 2003, and so on. The MATLAB after operation is carried out, obtained Huanggang city visitors quantity fitting results and errors are shown in Table 1, the predictive values of the next few years are shown in Table 2.

Table 1. Fitting value and error of visitors quantity in Huanggang City based on the neural network model.

year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
fitting results(million tons)	229.2	251.1	286	312.1	340	507.9	742.7	747.9	909.4	688.9
error (%)	-0.7	0.4	2.02	-2.37	0.64	-0.005	-0.003	-0.0002	-0.001	0.001

Table 2. Visitors quantity forecast value based on the neural network model in Huanggang City.

Year	2014	2015	2016	2017	2018	2025
Forecast value (million tons)	1103	1227.9	1360.3	1500.4	1800.6	2412.5

F. *Time series regression analysis forecast*

In this paper, by means of SPSS software, to fit Huanggang city past years visitors quantity, use regression analysis method based on time series, get a regression analysis equation such as type 9,

Huanggang city visitors quantity fitting value and the error value is in Table 3, the predictive value is in Table 4.

$$y = 3.7919x^2 + 7.3523x + 139.49 \quad (9)$$

Table 3. Visitors quantity fitting value and the error value based on regression analysis forecast method in Huanggang city.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Fitting value (million ton)	169.4	195.7	229.6	271.1	320.1	376.8	440.9	512.8	592.2	679.2	773.8	875.9
errors (%)	9.1	-15.6	-0.5	8.4	14.2	17.8	30.5	0.9	-20.2	-9.2	-14.9	27.2

Table 4. Visitors quantity forecast value based on regression analysis forecast method in Huanggang city.

Year	2014	2015	2016	2017	2018	2025
Forecast value (million tons)	800	952.4	1050	1171.9	1966.1	2894

In which $x=1, 1, 2, \dots, 18$ (calculate the data represents 2000 year of Huanggang city visitors quantity, and so on.

G. *Grey prediction*

Considering the increase of original series dimension will affects the prediction results, this paper selects visitors quantity data in 2005 – 2012 Huanggang city as the raw data, in order to improve the prediction precision, take power function, exponential function model to preprocess the original data. That is the original series were multiplied by 10^{-3} then a^{-x^m} ($a > 1, m > 1$),

a^{-x^m} ($a > 1, m > 1$) transform, By calculations are comparison in the MATLAB, found that when $\frac{1}{a} = 0.8, m=1.05$, the accuracy is in the maximum, then set up GM (1,1) model, with the aid of the MATLAB analysis, Huanggang city visitors quantity data time response model type see type 10.

$$\hat{x}^{(1)}(t+1) = -45.5627e^{-0.020665t} + 46.5056 \quad (10)$$

Then exponential transform the fitting values and reduction obtained fitting values shown in Table 5.

Table 5. Fitting value and the error value of visitors quantity in Huanggang city based on Grey Theory.

year	2005	2006	2007	2008	2009	2010	2011	2012
Fitting value (million ton)	280.35	333.89	426.18	517.90	608.73	698.75	788.53	877.6
errors (%)	0	4.4	26.1	1.9	-18	-6.5	-13.3	27.4

Upon examination, the ratio of variance C was 0.4410, a small probability of error is 0.875, compare level reference table of grey forecasting accuracy (Table 6), the results are qualified, show that the

impact of prediction model can be accepted, and then use the time response model predicting visitors quantity value of Huanggang city in the next few years in Table 7.

Table 6. Grade reference table - testing of the accuracy.

Grade	P	C
Good(Class A)	>0.95	<0.35
Qualified	>0.8	<0.5
General	>0.7	<0.65
Unqualified	<=0.7	>=0.65

H. *Based on neural network combined prediction*

Use fitting value got in front of three kinds of prediction method as input samples, namely fitting the fitting value and the actual value obtained from

2005 in three methods, followed by analogy, sample number is 2005 ~2012, total data sets is 8, use the corresponding to the actual value as the output

samples, training the network learning, Amplify neural network to identify the weight of every predict method, reduce subjective factors in the prediction process. In network design, using the input layer is 3, the output layer is 1, the training samples is 8 group,

neurons is 5, get new fitting value in Table 8. Finally, prediction values got from three prediction methods as input samples, through the network to get the next few years Huanggang city visitors quantity forecasting values (Table 9).

Table 7. Visitors quantity forecast value of Huanggang city based on Grey Theory.

Year	2014	2015	2016	2017	2018	2025
Forecast value(million ton)	1054.2	1142.0	1229.4	1316.6	1576.4	2004.7

Table 8. Value of neural networks.

year	2005	2006	2007	2008	2009	2010	2011	2012
Fitting value (million ton)	282	278	298	504	734	742	913	694
errors (%)	0.57	-13.1	-11.9	-0.74	-1.25	-0.77	0.34	0.68

Table 9. Visitors quantity estimated value and the error value in Huanggang city based on combined prediction.

year	2014	2015	2016	2017	2018	2025
Prediction value(million ton)	1012.5	1243.1	1286.4	1417.8	1777.5	2355.7

According to the above results analysis and the deviation of fitting value and the actual value we can see that neural network combination forecast method is the most accurate, which shows that the method is feasible and effective results and accuracy, can be used for Huanggang city visitors prediction.

IV. DISCUSSION & CONCLUSIONS

Accurate prediction of visitors quantity can make tourism in the fierce market competition to reduce the risk, formulate reasonable tourism policy and launch a reasonable tourism product. This paper attempts to use the gray prediction model and the neural network prediction model of combined prediction. Finally, through an example to illustrate the method can effectively improve the prediction accuracy, reduce the prediction error. Based on the grey theory and neural network visitors quantity combined prediction model is proposed, extract the useful part of each single prediction model, But implementation of the combined prediction model is affected by many factors, how to improve the neural network's learning effect, determine the comprehensive prediction of nonlinear function and how to avoid every single prediction model provides a valid information, are in need of further study. Moreover, the method for short-term prediction is more effective and high precision, but for long term prediction is not necessarily to achieve satisfactory results, this is the learning direction in the study of combined prediction method to the direction of learning in the future.

ACKNOWLEDGEMENTS

This work was supported by humanities and social science research base project 2012CR007 of Hubei Provincial Education Department, "Dabie Mountain tourism core competitiveness and regional economic coordinated development research." Also was by supported by Project D20152901 of Science and technology project of Hubei Provincial Education Department.

REFERENCES

- [1] K. C Luk, J. E. Bal, and L. A. Sharma, "A study of optimalmodellag and spatial inputs to artificialneuralnetwork for rainfall forecas-ting," *Journal ofHydrology*, no. 227, pp. 56-65, 2000.
- [2] Guibing Zhao, Yanfu Shi, Wenfeng Duan, and Huarui Yu, "Computing fractal dimension and the kolmogorov entropy fromchaotic time series," *Chinese Journal of ComputationPhysics*, vol. 16, no. 3, pp. 309-315, 1999.
- [3] Jinhu Lv, Jun'an Lu, and Shihua Chen, "Chaotic time seriesanalysis and it'sapplications," WuHan: Wuhan University Press, 2002.
- [4] Keng Chen and Botang Han, "A survey of state space re-construction of chaotic time series analysis," *ComputerScience*, vol. 32, no. 4, pp. 67-70, 2005,
- [5] Xinhui Wen and Minghao Niu, "A neural network based nonlinear combination forecasting method," *System engineering theory and Practice*, no. 12, 1994.
- [6] Qing Zhang, "Study on the Neural network based optimal combination forecasting method," *System engineering theory and Practice*, no. 9, 2001.
- [7] Tang Luo, "Optimization grey model," *Chinese Academy of Engineering*, no. 15, 2003.
- [8] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamicsdelaytimes and embedding windows," *Physica D*, no. 127, pp. 48-60, 1999.