The Frontiers of Society, Science and Technology
ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

Construction and Application of Forecast Model for Eco-Tourism Market Demand Based on BP Neural Network

Huize Chen

Xi'an Aeronautical University, Xi'an Shanxi 710077, China

ABSTRACT. *Now, analysis of the tourism market in the future period of time usually uses time series prediction and causal model prediction. However, due to the complicated tourism market environment, the small change of factor will affect the final result. Therefore, these methods can't predict the future market conditions of tourism very accurately. Since BP neural network is a mathematical algorithm for simulating the structure of the brain's neural network, it has a very good processing capability for such non-linear relationship problems, can sort out complicated and various factors and finish the model building systematically, and finally get the effective forecast result. Prediction of the tourism market is a typical multi-factor impact, and there are a lot of unknown factors of nonlinear relationship, so BP neural network is very suitable for analysis and processing of tourism market.*

KEYWORDS: *Eco-tourism market; BP neural network; Demand forecasting model*

1. Introduction

As the industrial revolution has driven the innovation and progress of various industries, and after decades of development, the level of economy and culture in the society has now reached a certain height (Thi N, et al. 2017) [1]. Now more and more people recognize the importance of a healthy ecological environment to human beings and pay more and more attention to ecological protection (Xiao-Jing M A, et al. 2017) [2]. However, the spiritual and cultural needs emerge after tourism has certain height of social development. The protection of ecological environment is also very obvious in the tourism industry, and different from the previous tourism development which only focuses on economic benefits. The development of ecotourism has been more and more in recent years due to the increasing demand for ecotourism (He X Q, et al. 2016) [3]. Although the development of ecotourism is popular, there are also obvious flaws. The lack of relevant experience in ecotourism management and the lack of detailed planning for tourism development have seriously hindered the development of the industry. Therefore, it is necessary to make relevant analysis on the ecotourism industry, predict the demand for tourism and the law of tourism, so as to make a better plan for the development of ecotourism.

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

2. Methodology

2.1 BP Neural Network Algorithm

Multiple neurons are interconnected to form a neural network, which can simulate the reaction of the biological nervous system to the real-life things, and people can also conduct targeted training through the sample of the external parameters. The neural network can memory the training without relying on other fixed algorithms, can summarize the best solution to deal with this kind of events from the training results and summarize the algorithm to deal with the calculation and understand the rules existing in the data (Zhokhova V, et al. 2016) [4]. It is because of this nature of neural network algorithms that makes it well-suited to dealing with things with complicated causal relationships and many unpredictable effects. Through the use of very rich data in event, BP neural network training is conducted, so that the statistical analysis of such events is realized, and the final result is obtained. BP neural network model is shown in figure 1.

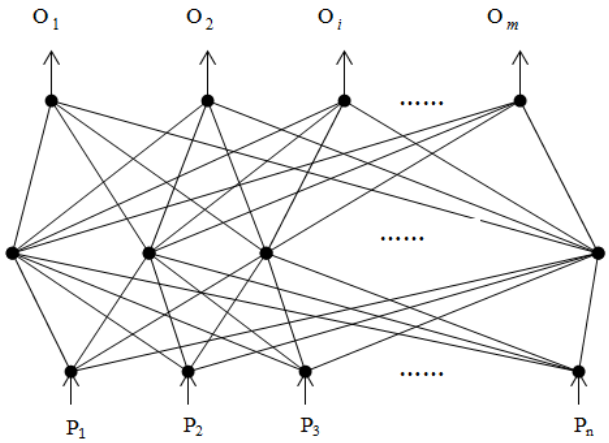

Figure.1 Bp Neural Network Model Diagram

The learning of BP neural network algorithm is composed of two steps, namely forward propagation of signal and error reverse propagation. The former deals with the event at the input layer, and then the next stage are the reverse propagation of error. The error is dealt with in a specific way and returned to the input layer again, the signal is equally distributed to each unit, thus the error of the unit is obtained, according to the error situation, the error correction scheme is obtained, and the process is repeated, so that the weights are corrected, this is the learning process of the neural network (Chen L, et al. 2016) [5]. This correction, or learning process will continue until the result of the calculation has reached an acceptable level or reaches the previously set number of studies. BP network is a network structure formed by

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

the interleaving of input layer and output layer and other units, this allows the input layer and the output layer to establish its nonlinear relationship and greatly expand its use in the field, so that the output result is no longer between -1 to 1. Specific steps are as follows: (1) initialize, the connection weights [*w*], [*v*] and the thresholds θ_i , r_t are randomly given; (2) the given input and output modes are used to calculate the output of the hidden layer and the output layer; (3) new connection weights and thresholds are calculated; (4) Returns the step by selecting the next input pattern (5) after repeated training, the error reaches the requirement.

First of all, the BP neural network model needs to be initialized. Each connection weights are given a random number in the interval (-1, 1), the error function *e* is set, the calculation precision and the maximum number of learning are given, the k-th input sample and the corresponding expected output are randomly selected.

$$d_0(k) = \left(d_1(k), d_2(k), ...d_q(k)\right) \tag{1}$$

$$x(k) = \left(x_1(k), x_2(k), ...x_n(k)\right) \tag{2}$$

The input and output of each neuron in the hidden layer and the partial derivatives $\delta_0(k)a$ of the neurons of the output layer by the error function using the network expected output and the actual output are calculated. The connection weight of the hidden layer to the output layer, the $\delta_0(k)$ of the output layer and the output of the hidden layer are used to calculate the partial derivative $\delta_h(k)$ of the error function for each neuron in the hidden layer. $\delta_0(k)$ of each neuron in the output layer and each neuron in the hidden layer are used to correct the connection weights $w_{h0}(k)$. $\delta_h(k)$ of each neuron in the hidden layer and each neuron in the input layer are used to correct the connection weights. Global error is calculated as follows:

$$E = \frac{1}{2m}\sum_{k=1}^{m}\sum_{o=1}^{q}(d_o(k) - y_o(k))^2 \tag{3}$$

The error is verified to see if it meets the planning requirements. The algorithm can be stopped when the error has been reduced to the planned range, or the number of training has reached a predetermined number of times. Otherwise, restarting the next round of algorithmic training is needed.

2.2 Gray Prediction Gm (1,1) Algorithm

It is necessary to define the gray forecast orderly so as to achieve the fundamental purpose of accurately predicting the unknown data. At present, the gray forecasting methods mainly include the following types: (1) gray time series forecasting; the algorithm model directly constructs the gray forecasting model based on the time series of the specific target to be predicted, and the model is applied to make accurate predictions for the unknown data at a certain time in the future. (2) Distortion prediction. The algorithm model mainly considers that the gray

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

prediction algorithm is prone to predict abnormal values in the process of prediction. Therefore, the model directly optimizes the distortion data in the gray prediction algorithm to predict accurately the unknown data at a certain point in the future. (3) System prediction: The algorithm model is mainly to predict the change of coordination relationship among many variables in the system through the establishment of the gray prediction model with a global concept. (4) Topology prediction: the algorithm mainly constructs the curve of the original data, based on the curve drawing, the mathematical formula is used to accurately determine the possible value at a particular time point, then, these different values are formed into an ordered time series, and then the model is established to predict the time of the value. The establishment steps of grey prediction GM (1, 1) model are as follows:

First of all, the original data needs to be processed. In order to weaken the impact of the randomness of the original time series on the prediction results during the data processing, before the gray prediction GM (1,1) model is established, its original time sequences are processed sequentially and columns are generated.

$$X^{(0)}=\left\{X^{(0)}(1),X^{(0)}(2),X^{(0)}(3),...X^{(0)}(n)\right\} \tag{4}$$

$$\lambda(t)=\frac{X^{(0)}(t-1)}{X^{(0)}(t)},t=2,3,\cdots,n \tag{5}$$

$X^{(0)}$={$X^{(0)}$(1), $X^{(0)}$ (2), $X^{(0)}$)(3),…$X^{(0)}$(n)} is the original data of an indicator to be predicted, the order ratio of a sequence of numbers is calculated as formula (5), the GM (1,1) model can be established and grey prediction can be carried out if most of the class ratios are in the coverage area $(e^{\frac{-2}{n+1}},e^{\frac{2}{n+1}})$. Otherwise, the data is properly pre-processed. The preprocessed data is smoothed into a three-point smoothing, which can be handled as follows:

$$X^{(0)}(t)=\left[X^{(0)}(t-1)+2X^{(0)}(t)+X^{(0)}(t+1)\right]/4 \tag{6}$$

$$X^{(0)}(n)=\left[X^{(0)}(n-1)+3X^{(0)}(n)\right]/4 \tag{7}$$

$$X^{(0)}(n)=\left[X^{(0)}(n-1)+3X^{(0)}(n)\right]/4 \tag{8}$$

After preprocessing, the data is accumulated and generated. That is, the first data of the original sequence is taken as the first data of the generated column, the second data of the original sequence is added to the first data of the original sequence, and as the second data of the generated column. According to this rule, generated columns can be obtained. According to $x^{(1)}(k)=\sum_{n=1}^{k}X^{(0)}(n)$, a new series can be obtained.

$$X^{(1)}=\left\{X^{(1)}(1),\ X^{(1)}(2),\ X^{(1)}(3),\ldots X^{(1)}(n)\right\} \tag{9}$$

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

The randomness of this newly formed series is greatly weakened compared to the original data, and the stability needs to be greatly increased. The differential equation is as follows:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \tag{10}$$

The a in the formula is the grey number, and the u is the endogenous control grey number.

If $\hat{\alpha}$ $Y_n = [X^{(0)}(2), X^{(0)}(3), \ldots, X^{(0)}(n)]^T$, $\hat{\alpha}$ is parameter vector to be estimated $\hat{\alpha} = \begin{pmatrix} a \\ u \end{pmatrix}$,

$$B = \begin{bmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \cdots & \cdots \\ -\frac{1}{2}(X^{(1)}(n-1) + X^{(1)}(n)) & 1 \end{bmatrix} \tag{11}$$

The model can be expressed as $Y_n = B\hat{\alpha}$ and $\hat{\alpha} = (B^T B)^{-1} B^T Y_n$ can be obtained by the least square method, so the discrete time response function of grey prediction can be obtained:

$$\hat{X}^{(1)}(t+1) = \left[X^{(0)}(1) - \frac{u}{a} \right] e^{-at} + \frac{u}{a}, t = 0, 1, 2..., n-1 \tag{11}$$

$\hat{X}^{(1)}(t+1)$ is accumulated predictive value of the obtained, and the prediction value is reduced as:

$$\hat{X}^{(0)}(t+1) = \hat{X}^{(1)}(t+1) - \hat{X}^{(1)}(t) \tag{12}$$

3. Result Analysis and Discussion

The BP neural network algorithm and gray prediction GM (1,1) algorithm were used to design a demand forecasting model of ecotourism market. Tourism demand refers to the natural beauty of a place that is attractive to some people who have the ability to travel and pay. These people are willing to spend some time in the area to travel and buy some travel products. The demand for travel markets in a region is often expressed in terms of the number of tourists that a region receives over a period of time. There are generally the following factors: (1) tourism resources and environment; (2) tourists spending power; (3) local consumer spending index; (4) service quality; (5) seasonal factors; (7) traffic factors, as shown in table 1.

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

Table1 Weights Index Tables at Different Levels

Means of transportation	Airplane	Train	Automobile
Proportion	84.6%	0.6%	14.8%

Based on the above analysis, as long as the relationship between these factors and the volume of tourists are analyzed, and the function between tourist volume and these factors is the established, demand forecast of tourism will be more reasonable and accurate.

Based on the above analysis of the six major factors affecting the number of tourists, the general model of the number of tourists on demand can be expressed as follows:

$$Q = F\left(l, m, c, z, s, j\right) \tag{13}$$

In this formula, the variable Q mainly represents the number of tourists who have demand in tourism. The variable l mainly represents the tourism resources and environment. The variable m mainly represents the consumption power of tourists. The variable c represents the local residents' consumption index, z indicates the quality of service, s indicates seasonal factor and j means traffic factor.

Taking into account the needs of the ecotourism market in the tourist area, there are many factors that affect the tourism industry. Therefore, the least squares criterion is used to establish the multiple linear regression model, and its linear form can be expressed as:

$$Q = Q_0 + \gamma_1 l + \gamma_2 m + \gamma_3 c + \gamma_4 z + \gamma_5 s + \gamma_6 j + \mu \tag{14}$$

Combined with the above analysis of the impact of the factors, it can be seen that the main influencing factors are tourists spending power, local consumer spending index, quality of service, and the decisive factors is tourism resources and environment, K indicates this kind of influencing factor, as shown in table 2. In order to simplify the calculation, the above linear model can be directly adjusted as follows:

$$y = k \bullet f(x_1, x_2, x_3) = k\beta_0 + k\beta_1 x_1 + k\beta_2 x_2 + k\beta_3 x_3 + k\varepsilon \tag{15}$$

Table 2 Seasonal Factors in the Demand of Eco-Tourism Market

Year	Visitors to the total y_1	Foreign population y_2	Per capita expenditure $x_1(t)$	Consumer price index $x_2(t)$ (%)	Service quality satisfactionx3(t) (%)
2008	1254.96	38.93	760.02	103.6	63.4
2009	1234.11	29.33	758.04	100.8	65.7
2010	1402.89	30.86	791.30	102.1	68.2
2011	1516.47	43.19	824.61	101.5	69.9

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

2012	1605.02	46.57	881.17	101.5	72.4
2013	1845.51	75.31	928.58	105.0	76.5
2014	2060.00	97.93	933.64	106.9	79.1
2015	2250.33	55.15	940.84	99.3	83.5
2016	2587.34	66.31	995.73	104.1	87.6

After using Matlab software to perform multiple linear processing on the data in the table, the following results can be obtained as shown in table 3:

Table 3 Results of Multiple Linear Processing on the Data

Regression coefficient	The estimate of regression coefficients	The confidence interval for regression coefficients
$k\beta_0$	-4106.9218	[-6541.2562 -1672.5874]
$k\beta_1$	-1.3802	[-4.0544 1.2940]
$k\beta_2$	18.3580	[-7.4003 44.1162]
$k\beta_3$	69.8281	[42.4262 97.2299]
R^2 =0.9917 F=198.3232 p =0.00001288<0.05		

So, the initial regression equation is obtained:

$$y = -4106.9218 - 1.3802x_1 + 18.3580x_2 + 69.8281x_3 \quad (16)$$

After obtaining the regression equation of the model, the reliability of the model needs to be tested, the value of measurement *R*, *F* and *P* are mainly used to judge whether the model is available or not. Evaluation of correlation coefficients R^2 : in general, a reliable prediction model requires that the value of the correlation coefficient be controlled between (0.8, 1). The closer the value of the relevant parameter to 1 means the independent variables of and the dependent variables in the regression equation have a strong correlation. The absolute value of R^2 in this article is 0.9917, indicating a strong linear correlation. F-test: when $F > F_{1-\alpha}(m, n-m-1)$, there is a significant linear correlation between the dependent variable y and independent variables x_1, x_2,…, x_m; otherwise, the linear correlation between the dependent variable *y* and the independent variables x_1, x_2,…, x_m is not significant. This example F=198.3232> 3.8625. P-value test: If $p < \alpha$ (α is the predetermined significant level), which indicates that there is a significant linear correlation between the dependent variable *y* and independent variables x_1, x_2,…, x_m. The output result of this example is p =0.000012888, which apparently satisfies $p < \alpha$ =0.05.

If the results obtained by the above three statistical methods show a high consistency, there is a clear linear relationship between the dependent variable *y* and the independent variable, and the regression model obtained from the linear relationship is accurate and effective. This regression equation can be directly used to predict the number of tourists, as shown in table 4:

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

Table 4 The Comparison between the Number of Tourists

Year	Actual value	Predictive value	Relative error
2008	1254.96	1173.09	-0.0652
2009	1234.11	1285.02	0.0413
2010	1402.89	1437.55	0.0247
2011	1516.47	1499.27	-0.0113
2012	1605.02	1595.78	-0.0058
2013	1845.51	1880.89	0.0192
2014	2060.00	2090.34	0.0147
2015	2250.33	2248.13	-0.0010
2016	2587.34	2546.78	-0.0157

The data in the above table is made into a line chart. As shown in figure 2, it is easy to see that the number of tourists predicted by the demand forecasting model designed in this paper is close to the actual number of tourists. The subtle errors are within the tolerance of mathematical calculations, which also shows that the model of market demand forecast for ecological tourism based on BP neural network has certain practical value.

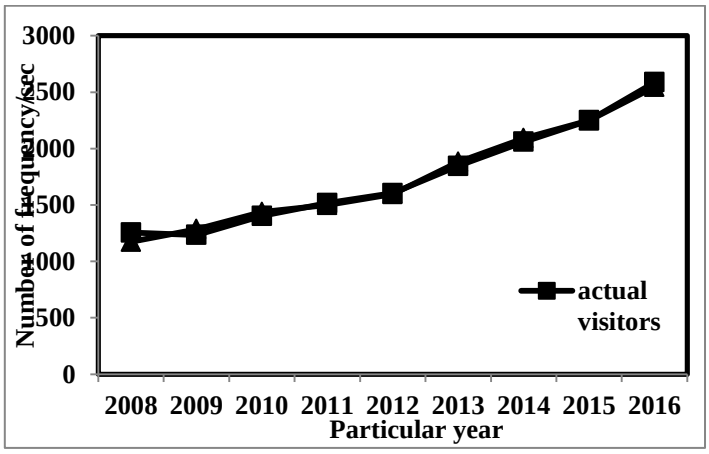

Figure.2 The Comparison between the Predicted Value and the Actual Value

4. Conclusions

Due to the progress of the algorithm and the increasingly powerful computational power of the computer, the algorithm of the machine is becoming more and more mature and practical, and can reduce people's workload in many aspects, which makes the machine algorithm applied in real life widely. With the increasing demand for eco-tourism, more and more eco-tourism is being developed. Although the development of ecotourism is popular, there are also obvious flaws.

ISSN 2616-7433 Vol. 1, Issue 6: 87-95, DOI: 10.25236/FSST.2019.010617

The lack of relevant experience in ecotourism management and the lack of detailed planning for tourism development have seriously hindered the development of the industry. Therefore, it is necessary to make relevant analysis on the ecotourism industry, predict the demand and law of tourism, and make a better plan for the development of ecotourism. Therefore, in this paper, BP neural network algorithm and gray forecasting GM (1,1) algorithm were used to design a demand forecasting model for ecotourism market, and explore how to make a good forecast of tourism demand through this demand forecasting model, so as to achieve effective optimization of tourism planning and promote the healthy development of eco-tourism purposes.

References

[1] Thi N, Wen Z (2017). Predictive Study of Eco-tourism Market Demand Based on Grey Model (1, N)--A Case Study of Phong Nha-Ke Bang, Quang Binh, Vietnam. Journal of Green Science & Technology, vol.11, no.8, pp.27-28.

[2] Xiao-Jing M A (2017). Ethnic eco-tourism: an effective model for protective development of ethnic tourism. Human Geography, vol.72, no.3, pp.961-980.

[3] He X Q, Liu Y (2016). Discussion on the development model and path of rural eco-tourism in Chongyi county. Guizhou Agricultural Sciences, vol.37, no.4, pp.179-181.

[4] Zhokhova V (2016). Model of competitiveness of eco-tourism product (by the example of tourism organizations in primorsky krai). Middle East Journal of Scientific Research, vol.18, no.3, pp.488-492.

[5] Chen L, Wang S H, Ding H F, et al (2016). The research on development strategy of eco-tourism on desert island. Economic Geography, vol.38, no.3, pp.211-216.