Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(3):1121-1125

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Tourist traffic prediction method based on the RBF neural network

Li-Juan Liu

School of Economics and Management, Yanshan University, Qinhuangdao, China

ABSTRACT

It is of great significance for scenic spots to establish an accurate prediction model which can reflect on tourist flow and has quantitative relation with other factors. For this reason, a tourist traffic prediction method is proposed based on the RBF neural network. This method has greatly improved work efficiency, provided guarantee for the stability of numerical tourist traffic prediction.

Key words: *Tourist Traffic Prediction; Neural Network; Radial Basis Function*

INTRODUCTION

According to statistics of WorldTravel and Tourism Council (WTTC), tourism economy of China will rank No. 3 in the whole world and growth in tourism demand will rank No. 1 worldwide by 2012 [1-2]. As tourism demand grows rapidly, the number of visitors at many hot scenic spots stays in a saturation condition, especially in peak tourist season, which not only proposes baptism for reception and accommodation ability of scenic spots as well as their hotels and guesthouses but also goes against protection of tourism resources in scenic spots. Currently, most of scenic spots in China have begun to pay attention to prediction about tourist flow. However, since informatization starts late and data sample about tourist flow with records is small, all predictions on tourist flow are based on emotional management experience and macro-forecast, and there is no clear mathematical model, there is huge deviation between predicted value of tourist flow and its real value, and it is impossible to establish quantitative relation between tourist flow and other factors, for example, historical tourist flow. Therefore, it is of great significance for scenic spots to establish an accurate prediction model which can reflect on tourist flow and has quantitative relation with other factors, for example, historical tourist flow.

Early prediction on tourist flow mainly focuses on econometric models [3-4], whose main feature is that tourism demand is predicted by analyzing factors affecting tourist flow. However, it is difficult, time-consuming and expensive to determine econometric models [5-6]. Traditional time series methods, such as exponential smoothing method, linear regression method, grey prediction method, ARMA method, ARIMA method and SARIMA method etc., pay attention to 'letting data to speak by themselves', and they are more appropriate for prediction on tourist flow. However, since they lack a learning process about data sample, it is difficult for them to realize complicated and nonlinear prediction on tourist flow.

RBF NEURAL NETWORK

RBF neural network is a traditional technique of multi-dimensional spatial interpolation, which has favorable global approximation capability and composed of an input layer, a hidden layer and an output layer. In addition, it is a neural network model proposed to overcome the two defects BP neural network has, i.e., neural network and slow rate of convergence, whose structure is shown in Figure 1.

As a form of primary function on the hidden layer, gaussian kernel function (GKF) has been used most commonly:

$$R_j(\boldsymbol{X}-\boldsymbol{c}_j)=\exp(-\left\|\boldsymbol{X}-\boldsymbol{c}_j\right\|^2/2\sigma_j^2),\ \ j=1,2,...,p$$

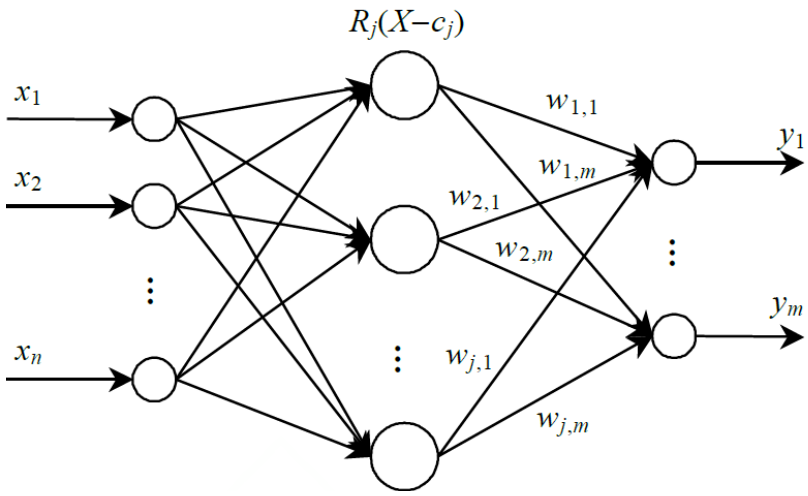

Fig. 1 RBF neural network

Where: $X = [x_1, x_2, \cdots, x_n]$ is n -dimensional input vector, c_j is the center of the j^{th} primary function and is a vector with the same dimension as X , σ_j acts as a generalized constant of the j^{th} nerve cell, i.e., variance of GKF; n and p refer to the number of nerve cells on the input layer and the hidden layer, respectively. After the hidden layer function has been decided, the relational expression between input and output of RBF network is

$$y_i = \sum_{j=1}^{p} w_{j,i} \exp(-\|x - c_j\|^2 / 2\sigma_j^2), \quad i = 1, 2, \ldots, m$$

Where m stands for the number of nerve cells on the output layer, y_i is output value of the i^{th} nerve cell on the output layer and $w_{j,i}$ represents value of connection weight between the j^{th} unit on the hidden layer and the i^{th} unit on the output layer. For establishment of RBF neural network structure, 3 parameters need be solved, i.e., c_j data centre of primary function, the variance σ_j , and $w_{j,i}$ the weight from the hidden layer to the output layer.

RBF neural network belongs to forward neural network, whose structure has adaptivity and whose output has not relationship with initial weight value. The basic thought that realizes RBF neural network uses RBF as 'foundation' of the hidden unit to constitute space of the hidden layer. The primary function in the hidden layer makes response to input signals partially, i.e., it converts input vector and transform lower-dimensional pattern data into higher-dimensional space to make nodes on the hidden layer generate large output. In this way, the problem of linear impartibility in low-dimensional space can be linear separability in higher-dimensional space. Compared with other kinds of forward network, RBF network is featured by simple structure, concise training, high rate of convergence, favorable partial approximation performance and few parameters that need be set up. Thus, it has been widely applied to the field of science, such as nonlinear optimization, time series prediction and pattern recognition etc.

PROPOSED PREDICTION METHOD

Artificial neural network can simulate nonlinear structure from input to output, which provides an effective tool for us to solve nonlinear form. Based on single hidden-layer neural network, Taylor puts forward quantile regression model about neural network:

$$Q_Y(\tau \mid \boldsymbol{X}) = f[\boldsymbol{X}, \boldsymbol{W}(\tau), \boldsymbol{V}(\tau)]$$

Where, $\boldsymbol{W}(\tau) = \{w_{jk}(\tau)\}_{j=1,2,\ldots,J;\ k=1,2,\ldots,K}$ is weight vector of connection between the input layer and the hidden layer; $\boldsymbol{V}(\tau) = \{v_j\}_{j=1,2,\ldots,J}$ stands for weight vector of connection between the hidden layer and the output layer; K is the number of nodes on the hidden layer; and $f[\boldsymbol{X}, \boldsymbol{W}(\tau), \boldsymbol{V}(\tau)]$ is a nonlinear function formed by combination of $\boldsymbol{W}(\tau)$ and $\boldsymbol{V}(\tau)$, whose expression is:

$$f[\boldsymbol{X},\boldsymbol{W}(\tau),\boldsymbol{V}(\tau)]=g_2\{\sum_{j=1}^{J}v_j(\tau)g_1[\sum_{k=1}^{K}w_{jk}(\tau)x_{ki}]\}$$

It reflects a nonlinear structure from the explaining variable $\boldsymbol{X}$ to the response variable Y. Here, neural network selects RBF neural network structure. Nodes of the input layer directly spread signals to the hidden layer. For nodes on the hidden layer, radial action function g_1(-) is constituted by gaussian kernel function expression (1) etc. Nodes on the output layer can be converted by the linear function g_2(-). Especially, when g_1(-) and g_2(-) are equivalent functions, Quantile regression modular of neural network is a linear quantile regression model. In quantile regression modular of neural network, estimation on $W(\tau)$ and $V(\tau)$ can be converted into solving optimization problems:

$$\begin{aligned}
&\min_{\boldsymbol{W},\boldsymbol{V}}\{\sum_{i=1}^{N}\rho_\tau[Y_i-f(\boldsymbol{X}_i,\boldsymbol{W},\boldsymbol{V})]+\lambda_1\sum_{j,i}w_{ji}^2+\\
&\quad\lambda_2\sum_i v_i^2\}=\min_{\boldsymbol{W},\boldsymbol{V}}(\sum_{i|Y_i\geq f(\boldsymbol{X}_i,\boldsymbol{W},\boldsymbol{V})}\tau\left|Y_i-\right.\\
&\quad\left.f(\boldsymbol{X}_i,\boldsymbol{W},\boldsymbol{V})\right|+\sum_{i|Y_i<f(\boldsymbol{X}_i,\boldsymbol{W},\boldsymbol{V})}(1-\tau)\left|Y_i-\right.\\
&\quad\left.f(\boldsymbol{X}_i,\boldsymbol{W},\boldsymbol{V})\right|+\lambda_1\sum_{j,i}w_{ji}^2+\lambda_2\sum_i v_i^2)
\end{aligned}$$

Where, λ_1 and λ_2 are penalty parameters to prevent network structure from falling into a state of overfitting. Donaldson et al. put forward cross validation method which is used to determine optimal values of penalty parameters λ_1 and λ_2 as well as K the number of nodes on the hidden layer. Standard gradient optimum algorithm can be used to obtain estimation on parameter vectors $W(\tau)$ and $V(\tau)$. After obtaining parameter estimation vectors $W(\tau)$ and $V(\tau)$, conditional quantile estimation on Y can be obtained.

$$\hat{Q}_Y(\tau\mid\boldsymbol{X})=f(\boldsymbol{X},\hat{\boldsymbol{W}}(\tau),\hat{\boldsymbol{V}}(\tau))$$

After conditional quantile estimation has been achieved, we may predict conditional density $\hat{P}_Y(\tau\mid\boldsymbol{X})$. When value of λ is taken continuously during the interval (0, 1), quantile curve Q is distribution function curve F.

RESULTS AND DISCUSSION

In order to verify rationality of the model and algorithm, this paper regards data about Huashan Mountain a famous 5A scenic spot in China as samples. The sample interval involves monthly data about tourist flow from 2010 to 2013. In accordance with needs of the model, dataset is divided into two parts, i.e., training set and training set. In detail, monthly data about tourist flow from 2010 to 2012 belong to the training set (36 months), while monthly data about tourist flow in 2013 (12 months) constitute the test set.

Since monthly tourist flow is large, normalization method is used to preprocess data to improve prediction accuracy, i.e., by $\frac{x_i-x_{\min}}{x_{\max}-x_{\min}}\times 0.7+0.15$, where $x_{\max}$ and $x_{\min}$ refer to maximum and minimum in the dataset.

Situations about prediction on the training set and the test set of the mode after parameters have been optimized are shown in Figure 2 where predicted values of the training set (36) and predicted values of the testing set (12) reflect fitting capability and predictive ability of the model, respectively. Besides, it can be found that fitting capability and predictive ability of RBF model are favorable and the fitting capability is superior to the prediction ability (average error of fitted value is 11%, while that of the predicted value is 12%).

In order to improve the predictive ability of the model, it is necessary to carry out seasonal adjustment for RBF model. According to the sequence chart about actual values (black line) of tourist flow in Figure 2, it is shown that the tourist flow of each year is featured by ‘three peaks and two valleys’ and periodicity and seasonality are obvious. Therefore, length of seasonal index number is set as 12. In accordance with 36 predicted valued obtained by fitting of the training set from 2010 to 2013, formulas are used to calculate seasonal index number of each month and predicted value after final adjustment (RBF values in Figure 3).

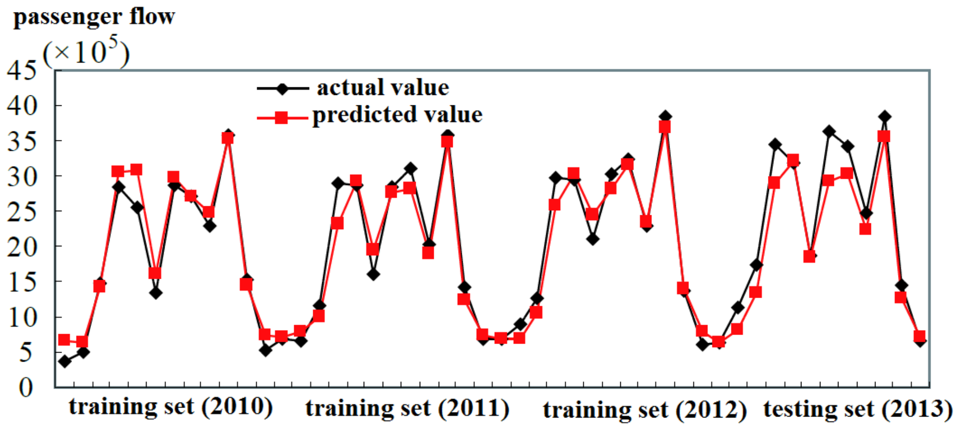

Fig. 2 Predicted Value of RBF Model

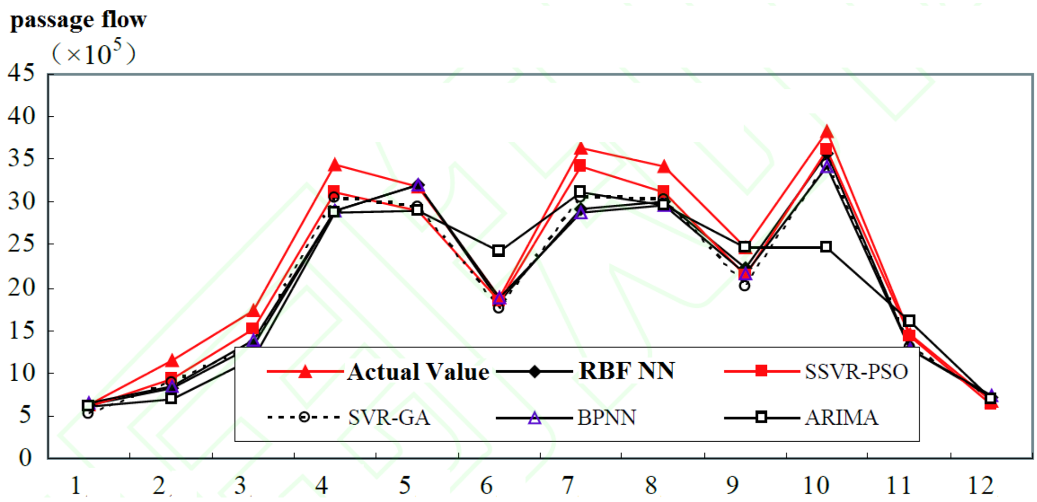

Fig. 3 Comparison about predicted values of each month by different methods

Among these seasonal index numbers, it indicates average predicted value of RBF model in the very month is lower than actual value when the seasonal index number is larger than 1, and the higher the seasonal index number is, the lower the average predicted value will be compared with the actual value. As a result, this may make scenic area management departments encounter situations, such as, insufficient manage resources. On the contrary, it implies that the average predicted value of RBF model in the very month is higher than the actual value when the seasonal index number is lower than 1, and the lower the seasonal index number is, the higher the average predicted value will be compared with the actual value. For scenic spots, the average value which is too high may leave management resources like manpower, financial resources and materials of scenic spots as well as result in waste. Via adjustment of seasonal index numbers, the predicted values of Huangshan Mountain Scenic Spot, which are lower or higher, are corrected timely and correctly so that final predicted values can approach actual ones to a larger extent, prediction accuracy can be improved largely and improper use of the scenic spot's management resources can be avoided.

Compare SSVR-PSO model with RBF, SVR-GA, BPNN and ARIMA to verify its accuracy and effectiveness. Details about the comparison are shown in Figure 2. Standards for evaluation on each model's advantages and disadvantages is mainly realized by R the correlation coefficient of the training set and the test set, MSE and average error (especially MSE and the correlation coefficient R of the test set). In detail, R stands for degree of relevance between predicted value and actual value, and MSE means the degree of deviation between predicted value and actual value. The closer the correlation coefficient R is to 1, the higher the degree of relevance between predicted value and actual value will be; the lower the value of MSE is, the smaller the deviation between predicted value and actual value will be. The predicted value obtained at this time is the most accurate.

CONCLUSION

This paper takes monthly data about tourist flow of Huashan Mountain a famous 5A scenic spot in China for example. Based on the situation that data about tourist flow have features like seasonality, nonlinear nature and small sample, it combines seasonal adjustment, and proposes a novel RBF model. It provides an effective and new method for prediction on tourist flow. This is quite important for development planning of the whole tourism industry because it enables scenic spots to cope with the phenomenon that jam may be caused when the number of tourists is too large in peak tourist season and during holidays and festivals sufficiently in advance as well as offers tourism management debarments important reference frames to make scientific decisions.

REFERENCES

[1] Lin C; Hu J; Kong XZ, *Chinese Journa*., **2012**, 35(1), 1-15
[2] Hu CP; Deng SL, *J. INTELL*., **2006**, 24(3), 321-325
[3] L Dudek, *ACM Trans. Comput. Hum. Interact*., **2006**, 18(1), 1-30
[4] M Dorigo; LM Gambardella, *IEEE T. EVOLUT. COMPUT*., **1997**, 1(1), 53-66
[5] LM Gambardella; ED Tailard; M Dorigo, **1999**. *J. OPER. RES. SOC*., **1999**, 50(2), 1167-1176
[6] D Costa; A Hertz, *J. OPER. RES. SOC*., **1997**, 48, 295-305