A New Combining Prediction Method of Visitor Numbers at Shanghai Expo

Shusheng Li, Rong Xie, Li Song, Xiaokang Yang, Wenjun Zhang
Institute of Image Communication and Information Processing
Shanghai Digital Media Processing and Transmission Key Lab
Shanghai Jiaotong University, Shanghai, China
lishusheng2008@gmail.com, {xierong, song_li,xkyang,zhangwenjun@sjtu.edu.cn}

Abstract. Forecast of visitor numbers to the large-scale activities is the key issue of collective behaviors analysis and control. At present, forecasting visitor numbers is mainly based on traditional research approach or sole artificial neural network technology. Recent study results show that combining forecast model approach enjoys more precise forecast than monomial forecast approach. In this paper, a new forecast approach based on inflexion point was proposed. Then, we combined BP neural network and the inflexion approach to make comprehensive analysis and to predict visitor numbers to Shanghai Expo per day. Experimental results indicate that the proposed combining approach is feasible and effective in forecast of the visitor numbers, and is more precise in terms of monomial forecast method. Respectively, the average relative error of combining model is 0.1085, 0.1177, 0.1875 less than that of "inflexion" model, BP model and ARIMA model.

Keywords: forecasting; BP model; "inflexion" model; visitor numbers; combining model

1. Introduction

The World Expo 2010 Shanghai China was the most expensive Expo in the history of the world's fairs. It had the largest number of countries participating and was the largest World's Fair site ever at 5.28 square km. By the end of the Expo, over 73 million people had visited and 250 countries and international organizations had participated. On October 16, 2010, the Expo set a single-day record of having over 1.03 million visitors enter the exhibition that day[1]. The successful prediction of visitor numbers will be most helpful for the organizers to reduce the risk of mass incident such as stampede and infectious diseases, and to take corresponding prevention measures to effectively prevent this type of cases.

The number of visitors to Shanghai Expo per day was affected by many factors which had complicated relationship. And it is difficult to predict the visitor numbers successfully by using single model, which cannot take into account both the linear and the non-linear changes. The essence of the combining forecast is that it sees the variety of individual forecasts as fragments with different information, scatters the uncertainty of the individual forecast and reduces the overall uncertainty through information integration, and improves the prediction accuracy [2].

The prediction of visitor numbers can be regarded as time series prediction. The time series prediction is the prediction of values basing on the previous values and the current value of the time series [3]. Many forecasting methods are available, but most of methods have their own merits and limitations, such as structure-based model method, classical statistical approaches and artificial intelligence techniques [4, 5].

BP model and ARIMA model are very practical forecasting methods in the time series prediction field. However, ARIMA model is only suitable for linear prediction and BP model is suitable for nonlinear prediction. To predict the number of visitors to Shanghai Expo per day more accurately, we need to combine the linear prediction with the nonlinear prediction. In this paper, we proposed a new model called the "inflexion" model for the linear prediction of the visitor numbers. And also a new prediction method combining the "inflexion" model and the BP neural network model was proposed to forecast the visitor numbers to Shanghai Expo. The "inflexion" model describes the linear relationship of the historical data, and the BP neural network model simulates the nonlinear law of the data. In addition, we compared the proposed combining model to the "inflexion" model, the ARIMA model and the BP model. These four models were constructed on the same of training data, which come from the statistical data of visitor numbers to Shanghai Expo from May 1 to September 30 in 2010. And the visitor numbers to Shanghai Expo from October 1 to October 31 in 2010 were predicted using these four established models. The experimental results show the prediction performance of the proposed combining method based on inflexion model and BP model is better than that of the other models. Respectively, the average relative error is 0.1085, 0.1177, 0.1875 less than that of inflexion model, BP model and ARIMA model.

The rest of paper is organized as follows: section 2 introduces "inflexion" prediction model, BP prediction model is described in section 3 and combining model is proposed in section 4. The experiments and results analysis is shown in section 5. A conclusion is drawn in section 6.

2. "Inflexion" Prediction Model

A. Characteristics of the Data.

We collected the statistical data of visitors at Shanghai Expo from May 1 to September 30 in 2010. We recorded the data per day. The data included 177 sampling points (one sample per 5 minutes) from 08:20:00 to 23:00:00. Fig.1 shows an example of data on June 28.

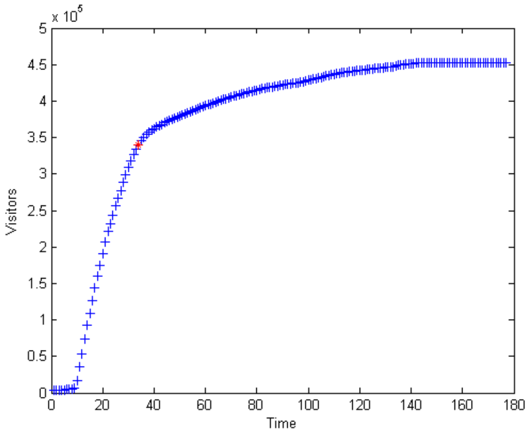

Fig.1 Example of data

From Fig. 1, we can see that the curve appears a turning point (the red point) which we call inflexion point. Through our research, we found that the inflexion point usually appeared before 12 o'clock and some degree of linear relationship existed between visitor numbers and the inflexion point(it is shown in Table 1). The value of inflexion point was bigger, the total number of visitors was bigger; the time of inflexion point was later, the total number of visitors was bigger. We can use inflexion point to predict the number of visitors every day.

B. How to Get Inflexion Point.

For one day, we got 177 sampling points (one sample per 5 minutes) from 08:20:00 to 23:00:00. We define the set

$X=\{x_1,x_2,...,x_{177}\}$. (1)

Then we calculate increment speed

$y_i=x_i+1-x_i, i=1,2,\ldots,176$. (2)

Firstly, we calculate the highest speed

$y_{max}=\max\{y_1,y_2,\ldots,y_{176}\}$. (3)

Secondly, we define the threshold speed

$y_{thr}=ay_{max}$. (4)

where a is a constant, $0\le a\le 1$. In this paper, a=0.4.

Thirdly, we define the acceleration threshold

$y_{a\text{-thr}}=by_{max}$. (5)

where b is a constant, $0\le b\le 1$. In this paper, b=0.05.

Finally, if $y_i-y_{i+1}>y_{a\text{-thr}}$ and $y_{i+1}<y_{thr}$, we will get the value of inflexion point: x_{i+1},and also know that inflexion point appeared at i+1.

C. How to Establish " Inflexion" Forecast Model

In this part, we will introduce how to establish "inflexion" forecast model. We calculated inflexion points from May 1 to September 30 in 2010. We define the set of value of inflexion points

$G=\{g_1,g_2,...,g_{153}\}$. (6)

Where g_j represents the value of inflexion point: x_{i+1}. We also define the set of time of inflexion points

$T=\{t_j=i+1 \mid g_j=x_{i+1}, j=1,2,...,153; i=1,2,...176\}$. (7)

We use N to represent the number of tourists every day. And then, we analyze the correlation of N, G and T, which can be shown in Table 1.

Table 1 depicts that the Pearson Correlation Coefficient between g and N is 0.887, and that between t and N is 0.586. On the 0.01 levels (bilateral), the relationship between g and N is significant correlation, so is that between t and N.

Finally we make use of G and T to establish " inflexion" forecast model

$N_j=kg_j+mt_j+b$. (8)

where k, m and b are undetermined parameters.

Tab.1 correlation

Variable		g	t	N
g	Pearson correlation	1	.842**	.887**
	Significant (bilateral)		.000	.000
t	Pearson correlation	.842**	1	.586**
	Significant (bilateral)	.000		.000
N	Pearson correlation	.887**	.586**	1
	Significant (bilateral)	.000	.000	

**. On 0.01 levels (bilateral) significant correlation

D. How to Use “Inflexion” Forecast Model to Predict Visitor Numbers

In order to predict N_p for one day p, we get the inflexion point (g_p, t_p) and then use Eq. 13 to get N_p.

3. Method of BP Forecast Model

Typically, the open EXPO system included people, vehicle, weather, temperature and so on. So the system’s dynamics features can’t be described perfectly by deterministic linear function. In this point, neural network may help us. Artificial neural network’s general expression is

$$y = \sum_{i=1}^{n} w_i x_i + b \tag{9}$$

where x_i(i=1,2,…,n) is the input value; w_i(i=1,2,…,n) is the weight; b is the threshold; y is the output value. If neural network has n input neurons, m output neurons and p hidden neurons, the output of neuron of hidden layer is

$$h_j = \sigma(\sum_{i=1}^{n} w_{ij} x_i + b_j),\ j = 1, 2, ..., p \tag{10}$$

and the output of neuron of output layer is

$$y_k = f(\sum_{j=1}^{p} w_{jk} h_j + b_k), k = 1, 2, ..., m \tag{11}$$

Usually, in the neural network, we use S type activation function

$$\sigma(x) = \frac{1}{1 + e^{-x/r}} \tag{12}$$

where r is the S type parameter of adjusting the form of activation function.

We use BP network to realize the prediction of visitor numbers to Shanghai Expo. BP neural network is the multi-layer feed-forward error back propagation neural network, usually contains input layer, output layer and several hidden layers. Each layer includes several nodes that represent neurons. Upper node and lower node are connected by the weight, and nodes between layers are connected with the form of whole internet, and there is no connection between nodes within each layer. Fig. 2 shows a typical BP neural network [7] which is a three layers structure network with a hidden layer.

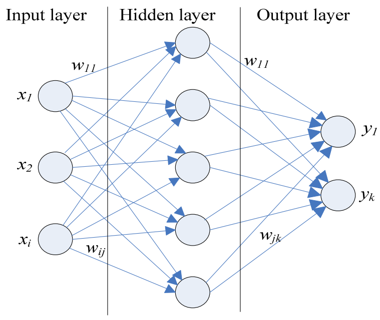

Fig.2 BP network structure.

Approximately, visitors prediction algorithm that based on BP network is concluded as follow:

1) Gathering and preprocessing original data to construct training samples.

2) Establishing network topology, defining the number of network layers, the number of input cells, the number of the hidden layer nodes, the number of output cells.

3) Designing learning algorithm to train neural network.

4) Evaluating network performance. If present accumulated errors are greater than given errors, then return 3); If present accumulated errors are less than given errors, then go into the step 5).

5) Inputting series of data, we can use the trained BP network to get the relevant prediction value.

4. Combining Prediction Model

Because the number of visitors to Shanghai Expo per day was often affected by many factors which had complex relationship including linear relationship and nonlinear law, using single forecast mode (BP neural network model or “inflexion” model) may make too many errors. So we firstly used the “inflexion” model to predict passenger flow, making its linear rule be contained in the predictive result of “inflexion” model. In this case, the nonlinear law contained forecast errors of “inflexion” model, and then we used the BP model to predict errors so that the nonlinear law was included in the prediction results. Lastly, we added both prediction results to get the predicted value of combining forecast model. Combining forecast model is shown in Fig. 3.

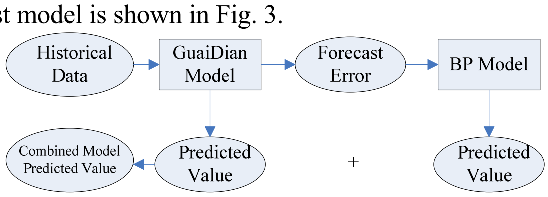

Fig.3 Combining forecast model principle diagram

5. Simulation and Comparison

A. "Inflexion" Forecast Model.

We used the data collected from May 1 to September 30 to establish the "inflexion" forecast model described in section 2. We got the " inflexion" forecast model

$$N_j = 1.588 g_j - 11460.583 t_j + 290421.105 \quad (13)$$

Fig. 4 shows that this model can very well fit historical data.

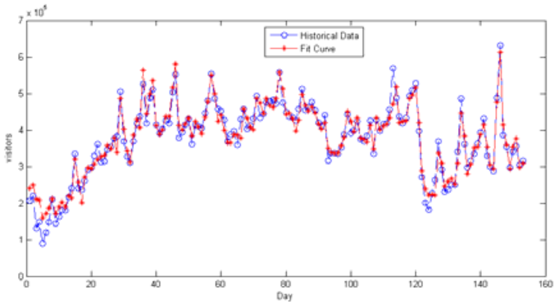

Fig.4 "Inflexion" model fit curve.

For October 1 to 31, we firstly got the inflexion point (g_j, t_j) depending on the sample points (one per 5 minutes) per day, then we made use of Eq. 13 to calculate visitor numbers that very day. At the same time, we used ARIMA model to predict visitor numbers. Both of the results are shown in Fig.5.

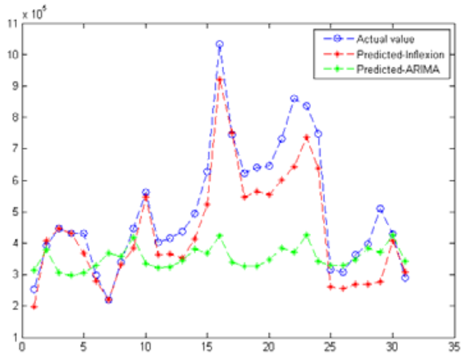

Fig.5 Comparison results using "inflexion" model and ARIMA model.

From Fig. 5, we can see that the predicted result of " inflexion" model is better than ARIMA model.

B. BP Forecast Model

We used the same data to establish the BP neural network, so the capacity of the training sample of BP neural network was 153. The input layer included 2 neurons, while the output layer included only one neuron (N-the number of visitors the very day). The BP neural network had only one hidden layer including 5 neurons that we set depending on the rule $2n+1(n=2)$. Before calculation, data should be normalized to the interval [0, 1]. The network structure used BP neural network structure with 2-5-1 based on LM algorithm. The maximum number of training epochs was 5000 and the network convergence error was 0.0005. We choose "tansig" as transfer function. After 34 times epochs, we could see the error approximation from Fig. 6. The predictive value of BP model is shown in Fig. 7.

C. Combining Forecast Model.

Firstly, we fitted the number of visitors from May 1 to September 30 using "inflexion" forecast model. At the same time, we got prediction errors. Secondly, we used these errors as training samples to establish BP neutral network with 2-5-1 based on LM algorithm. Thirdly, we predicted the number of visitors from October 1 to October 31 using the "inflexion" model. By this step, we predicted the linear law. And then, we used BP neural network model to predict errors. By BP model, we predicted the nonlinear part. Lastly, we added both of the two parts to predict visitor numbers from October 1 to October 31. Fig.7 shows that the predicted value obtained by the combining model is the closest to the actual value, the second is "inflexion" model, and the maximum is BP neural network.

Table 2 shows that the average relative error of combining model is 0.0167. It is 0.1085 less than that of inflexion model , is 0.1177 less than that of BP model and is 0.1875 less than that of ARIMA model.

6. Conclusions

In this paper, we presented procedures of three methods, including a new method "inflexion" model we proposed, BP model and combining model we also proposed. Simulation comparison of forecasting performances shows that the predicted value obtained by the combining model is the best closest to the actual value, the second is "inflexion" model, the third is BP neural network and the maximum is ARIMA model . From this, we can think that the combining model proposed in this paper gives a better method to solve the forecasting problem of visitor numbers at Shanghai Expo.

The combining model proposed in this paper can also be used in the large-scale activities. However visitor numbers to the large-scale activities is affected by many factors, such as weather factor, temperature factor, etc., so that it is often difficult to obtain absolute accurate predictive value of visitor numbers. Thus, the combining model is still waiting for being tested by practice for long.

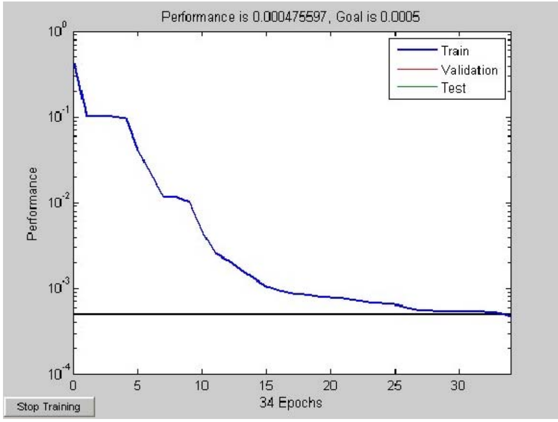

Fig.6 BP Network calculation epochs and the error chart.

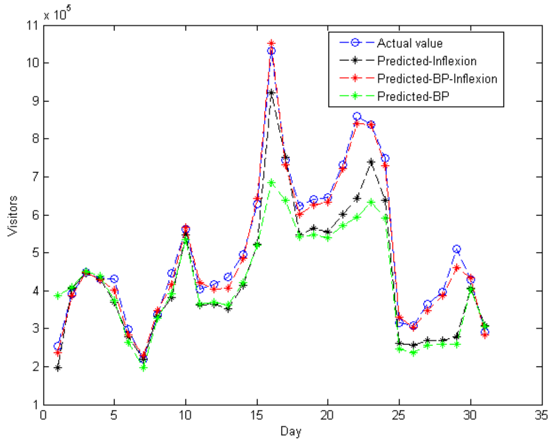

Fig.7 Comparison results using "inflexion" model, BP model and BP- Inflexion combining model.

Tab.2 average relative error

Method	combining	inflexion	BP	ARIMA
Relative error	0.0167	0.1252	0.1344	0.2042

7. Acknowledgments

This work is partially supported by National Basic Research Program of China (2010CB731400, 2010CB731401 and 2010CB731406) and NSFC project(NO. 60902020).

References

[1] "Shanghai World Expo wins worldwide applause". Xinhua. 31 October 2010. Retrieved 31 October 2010.

[2] Fen Chunshan, Wu Jiachun, Jiang Fu, "Combining Prediction of Oil Prices," Journal of University of Petroleum. China(Edition of Social Sciences), vol. 1, pp. 12–14, January 2004.

[3] I. Kaastra, M. Boyd, "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, pp. 215–236, 1996.

[4] Carse B.,Fogarty T. C., "Fast evolutionary learning of minimal radial basis function neural network using a genetic algorithm," Lecture Notes in Computer Science, vol. 1143, pp. 1–22, 1996.

[5] Fulcher G. E., Brown D. E., "A polynomial network for predicting temperature distributions," IEEE Transactions on Neural Network, vol. 5, No.3, pp. 372–379, 1999.

[6] Sheng Yanbo, "Forecast of Domestic Product Per Capita in Zhejiang Province Based on Combining Model of BP Neural Network and ARIMA Model,"Journal of Business reach, vol.23, pp. 49–50, April 2006.

[7] Fulcher G. E., Brown D. E., "A polynomial network for predicting temperature distributions," IEEE Transactions on Neural Network, vol. 5, No.3, pp. 372–379, 1999.