Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/080360

Forecasting Real-Time Tourist Arrivals using Hierarchical Cluster and Gaussian Fitting Algorithm: A Case Study of Jiuzhai Valley

*Lifei Yao, Ruimin Ma, Maozhu Jin**, *Peng Ge and Peiyu Ren*

Buisness School, Sichuan University, Chengdu 610064, China

Received: 1 Jul. 2013, Revised: 5 Nov. 2013, Accepted: 6 Nov. 2013
Published online: 1 May. 2014

Abstract: In the process of studying the number of tourist arrivals of each moment on every day, one of the key findings is a very good statistical regularity about tourist total arrivals. There exists high similarity among different scales of tourist arrivals, including the nearly same change trend. And on this basis this paper puts forward a new method of modeling and forecasting real-time tourist arrivals of each moment on every day. The forecasting process of the new proposed model is innovative. We consider the scales of tourist arrivals mainly, and model using hierarchical cluster and Gaussian fitting algorithm according to the different scales of tourist arrivals, thus predict real-time arrivals of the future a scale (the overall size of the known) through the existing scales. Finally take Jiuzhai Valley as an example to analysis, and experimental results show that the forecast method is effective.

Keywords: Tourist arrival; Gaussian fitting algorithm; Real-time forecast

1 Introduction

Over the past four decades, tourism has become one of the most rapidly growing industries [1]. Tourism contributes significantly to the economic growth of many countries and regions. The World Tourism Organization [2] reports that international tourist arrivals grew at a rate of 6% in 2007, reaching 900 million. Nowadays tourism plays a more and more important role in global economy. Thus accurate forecasts of future trends of tourism demand are particularly important to both tourism policymakers and tourism business practitioners [3].The increasing number of tourists brought problems for the scenic area management, the scenic area is becoming more and more crowded, especially holiday rush hour of tourists. But the distribution of tourists in scenic spots is not balanced in time and space dimension, that is to say, at the same moment some scenic nodes have too many tourists while others have few, and the unbalance and congestion will continue even expand without reasonable scheduling scheme. With the help of modern information technology, the scenic area managers hope to predict the number of tourist arrivals each moment on every day. Only by mastering the dynamic and real-time change rule can we carry out time and space optimization scheduling in time and make reasonable tourist route guidance. Thus scenic area can avoid or alleviate the congestion effectively and improve tourist satisfaction. Dynamic and real-time forecast is also a research emphasis in the future.

The research on forecasting of tourist arrivals began in the 1960s, and in recent years domestic and foreign scholars have used diverse methods in analyzing and forecasting tourist arrivals and have gained a rich research achievement. Pure time series approaches and models with explanatory variables are two common methods in the tourism forecast literatures. ARIMA forecasting [4], Innovations state space models for exponential smoothing [5,6,7] and Theta method [8] are classical models of time series approaches. Hyndman and Khandakar [9] proposed an automatic model selection algorithm in selecting an appropriate model order and improved the ARIMA model. Autoregressive distributed lag model (ADLM), Time varying parameter (TVP) model and Vector autoregressive (VAR) model are usual econometric methods. In order to improve the forecast accuracy, some scholars have tried to use intelligence artificial algorithms, such as the artificial neural network (ANN)

* Corresponding author e-mail: jinmaozhu@scu.edu.cn

method, the rough set approach and The fuzzy time-series method. Rob Law [10] built the first neural network model to forecast Japanese demand for travel to Hong Kong, showing that neural network model performed well in forecasting the tourism demand. Then neural network model had attracted lots of attention in forecasting such as [11,12,13,14]. Au and Law [15] deals with the classificatory analysis of imprecise, uncertain, or incomplete data by using the rough set approach, which pays much attention to the categorical variables such as demographic features and predicts tourism demand levels [16]. As far as the forecasting accuracy is concerned, the study shows that there is no single model that consistently outperforms other models in all situations [17]. Of course, combination methods among these forecast models above are common as well. Each forecasting method introduced into tourism forecast has both its original advantages and deficiencies. Different scholars used different methods to forecast tourist demand and test the merits of the model through the corresponding error evaluation standard. The conclusions among previous literature are a bit inconsistent because of the different research objects. Not a forecast method is absolutely the best one, only relatively.

From the analysis above, in addition to the most popular time-series and econometric models, a number of new techniques have emerged in the literatures [17]. The majority of these studies focus on the application of different techniques, both qualitative and quantitative, to model and forecast the demand for tourism in various destinations [18]. The forecast accuracy of new models is continuously improving at the same time. However, these models are generally limited to static forecasting of tourist arrivals and these studies used annual, quarterly or monthly historical data in modelling and forecasting tourism demand. It can be seen that the main data frequency in the previous literatures is still annual data. Studies used historical data at higher frequency such as daily, hour and minute are very few, in fact dynamic forecasting of tourist arrivals is of importance. Tourists' visiting activity begins since they arrive at the scenic entrance, tourists' arrival volume and arrival time are directly related to reasonable use and the optimized configuration of the scenic resources. At the same time, the capacity of scenic spots is limited, the tourist experience will be affected. In a actual scenic spot with only one entrance, one exit and many scenic nodes, tourists' motion law when entry into the scenic spot is similar to water's motion law in a complex pipe system, each node visits changes of each node is highly related to its upstream node, in essence, a node' tourists' number is the superposition of the results of all its upstream nodes' tourists which move to the node. The origin of this complex network transmission mechanism is the initial tourist arrivals volume to the scene entrance at each moment, which has a very important influence to the entire complex network system of scenic area. Therefore, modeling and forecasting the dynamic real-time laws of tourist arrival becomes the primary problem for posterior arranging line reasonably, making traffic plan and scheduling scenarios of travel tourism vehicles scientifically.

The rest of this paper is organized as follows. In Section 2, the proposed model for tourist forecasting is presented. Experimental design is presented in Section 3 concerning how numerical experiments are conducted. Section 4 presents and analyzes the experimental results. Finally, conclusions and future work are described in Section 5.

2 Methodology for tourist arrivals forecasting

Dynamic real-time forecasting of tourist arrivals requires more fine time scale and more higher frequent data such as daily, hour and minute level to analyze and research. Obviously, information technology plays a very important role, such as the strong support of the RFID information automatic acquisition system, RFID entrance guard system, etc. Along with the wide application of 'digital scenic spot' in Jiuzhaigou, the RFID tourist data acquisition system can obtain the location of the tourists and count data. Therefore the historical data includes the volume of tourist arrivals to the scenic spot entrance at each minute of every day, which provides the possibility of modeling and forecasting the dynamic and real-time variation pattern of daily tourist arrivals. In order to be able to describe how tourist volume of each node changes in the scenic system, it is very necessary that study the real-time changing law of tourist arrivals to the scenic entrance. Through modeling the historical data of different scales appropriately, we hope to achieve the goal of forecasting the number of tourist arrivals at each moment (minute) on some day in the future.

Fig. 1 shows the framework of the proposed model in this paper.

2.1 Data preprocessing

The arrival time of tourists to a scene entrance every day concerns about season, weather, accommodation, travel mode, age of the tourists, preference and so on. The arrival time of each tourist seems to be quite uncertain. When large amounts of tourists visit the same scene, what will happen? Observing the number of tourists to the entrance every minute on a day, we find it is a nonlinear and non-stationary time series, which seems that there is no regularity. However, if we get the integral of the number of tourists to the entrance every minute, that is to say, accumulating the number of tourists every minute, then we draw a graph about the accumulative value changing over time as seen in Figure 1. Let t denotes time (unit: minute), N_t denotes the number of tourists to the

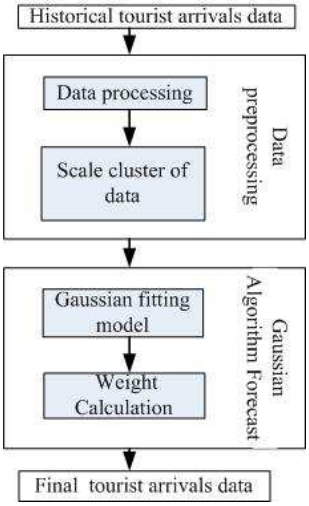

Fig. 1: Framework of the new proposed model

entrance at t, S_t denotes the total number of tourists arrivals to the entrance before , then

$$S_t = \sum_{i=1}^{m} N_t, t = 1,2,\cdots,m.$$

We can find out some statistical regularity as follows. There exists high similarity among the graphs of different scales, including the nearly same change trend. Besides, Figure 2 shows that growth rate of tourists to the entrance increases as the time increases in the beginning stage. And after a period of growth period, the growth rate gradually slows down to zero, and finally total tourists approach to its certain limit under different scales. In order to further study the problem, this paper introduces a much-needed boost in tourism growth 'speed' concept and related variables, on the basis of which is divided into three stages, trying to build tourism based on the theory of stage forecasting theory model, and then constructed stage forecast model and the corresponding mathematical statistics regression analysis method, and predicted the visits and the sample of model analysis and testing.

2.2 Scale cluster of data

Hierarchical Cluster Analysis is one of the most widely applied methods in clustering analysis [19]. Its basic principle is stated as follows. First we make certain amount of samples (or variables) into each class itself, and then find out the two classes with the mostly close nature to combine them into a new class. Next calculate the distance between all classes in the new class division, then merge the two classes with the mostly close nature. Repeat this process until all the samples (or variables) are merged into specified classes.

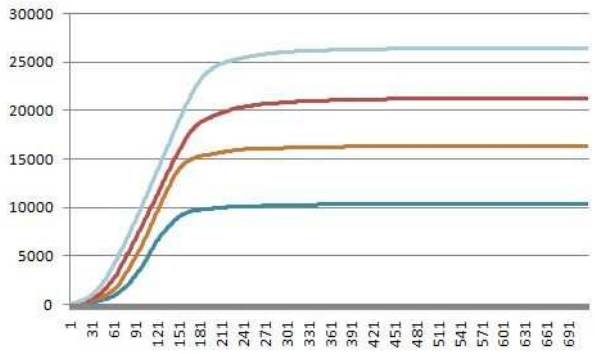

Fig. 2: Different dates time series of accumulative value

During hierarchical clustering process, the merger between class and class is mainly based on the distance between the two classes. Hierarchical Clustering procedure is shown as follows.

Step1:Give out a set Z with N samples, $s = Z_1, Z_2, \cdots, Z_n$

Step2:It will be gathered into K classes (K is prior given)

[1] $K = N$, $C_i = Z_i$, $i = i,2,\cdots,N$;

[2] If $k = K$, then END;

[3] Find out C_i and C_j with the minimum distance $d(C_i, C_j)$ among Z_i;

[4] Merge C_i and C_j into a new class C_i, and calculate the center of the new C_i;

[5] Remove C_i from $Z_i, k = k-1$, Go to [2].

In this paper, the hierarchical clustering is according to the tourist scale. We hope the distance of data scales in a merged class is close as much as possible, thus making the average distance of all data items among the merged class to be least. Therefore we use within-groups linkage method and define the distance.

$$d(C_i, C_j) = min_{Z_i \in C_i, Z_j \in C_j} \parallel Z_i - Z_j \parallel$$

2.3 Forecasting model based on Gaussian fitting algorithm

After the above hierarchical clustering, historical data of each day are divided into some discrete scales. This paper makes curve fitting according to different scales and gets a series of prediction models with the same structure but different parameters. During the actual fitting process, we does not require $y = f(x)$ strictly pass all the points (x_i, y_i), while only request the fitting error $D = f(x) - y_i$ of point x_i to be minimum according to a certain standard. It is a universal method that uses least squares approximation to find the best fitting curve.

This paper chose Gaussian function series as the basic fitting curve, that is to say, we set $y = F(x)$ based on the Gaussian function series [20]. Every Gaussian function is

all determined by three parameters, peak height A, peak position B and peak width C. The whole Gaussian function series is written as follows.

$$F(x) = \sum_{i=1}^{n} A_i * exp[-(\frac{x-B_i}{C_i})^2]$$

2.4 Calculate weight based on improved OWA operator

In 2.3, we have solved a series of forecasting models of different tourist arrivals scales. Obviously, these discrete scales do not include all scales of tourist arrivals because the tourist arrivals of every day may be different from others. When given the total number of tourist arrivals on a day, we hope to be able to forecast the number of tourist arrivals every minute. We can give these discrete scales proper weight to forecast the tourist arrivals of a certain actual scale. The idea of solving weight in this paper is based on OWA operator in literature [21], in which the different weights are aimed at different schemes or attributes. In order to get the needed weight about scales, this paper makes some corresponding improvement on the original basis.

The specific steps of calculating weights are given as follows according to this thought.

Step1: Calculate the offset distance d_{ij} between existing discrete scale i and target scale j, $d_{ij} = |v_i - v_j|, i = 1, 2, \cdots, m$, Where v_i represents the size of scale i;

Step2: Calculate weights;

$$\omega_i = \frac{d_{ij}^2}{\sum_{i=1}^{m} d_{ij}^2}$$

Step3: Sort the weights and scales in specific order;

$\omega = (\omega_1, \omega_2, \cdots, \omega_m)^m, \omega_1 > \omega_2 > \cdots > \omega_m$

$d = (d_{1j}, d_{2j}, \cdots, d_{mj}), d_{1j} < d_{2j} < \cdots < d_{mj}$

That is to say, which one is closer to the target scale, the weight is heavier.

Step4 Calculate the forecasting equation of target scale j.

$$f_i = \sum_{i=1}^{m} \omega_i f_i, i = 1, 2, \cdots, m$$

Where f_i represents the forecasting equation of tourist scale i.

3 Numerical example

To evaluate the performance of the proposed forecasting model, extensive experiments were conducted in terms of real scene data, which forecast the number of tourists every minute on a day basis.

3.1 Data resources and analysis

Real data were collected from Jiuzhaigou which is one of the most famous scenes in China. Take Jiuzhaigou scenic spot for example, we have collected real-time data of tourists' arrival time every day from May 2012 to August 2012 benefiting from RFID (Radio Frequency Identification Devices). The collected data almost lasts from 7 PM to 1 PM (unit of minute) every day, a total of 720 minutes. The play mode of Jiuzhaigou is 'travelling inside Valley, living outside Valley' and the tickets are only effective for one day, so tourists almost all arrive at Jiuzhaigou scenic spot before noon.

The scale of the tourists is seasonal, because the scales of similar dates are relatively close. In order to consider all kinds of tourist scales as much as possible in the process of forecasting, we should not only take the data of early dates as training data. Therefore we select them randomly as the training data to evaluate model parameters, and the rest are used to test the precision of model.

3.2 Hierarchical Cluster of tourist arrivals' scales

We make hierarchical clustering with the scales of historical daily data and the clustering result is listed below in Table 1.

The scales range from 10000 to 29000.The value of scale in Table 1 is arithmetic mean value of all historical data in the same layer and the number of tourist arrival every minute in the new scale is arithmetic mean value of the corresponding data of those days. Table 1 indicates that the data scales assigned in the same layer are similar. For example, data scales of the days in the first layer are all around 11000.

Line graphs of these different discrete scales all above are shown in the Figure 3, in which horizontal axis represents time t (unit minute), vertical axis represents the total number of tourist arrivals before the moment t. These curves from the bottom to the up mean that the scales increase in turn gradually.

Table 1 Result of Hierarchical Cluster (20 Layers)

Layer	1	2	3	4	5	6	7	8	9	10
Data	5.10	5.11	5.12	5.18	5.19	5.28	6.15	6.17	6.23	6.29
	5.14	5.13	5.20	5.26	5.25	5.29	6.16	7.5	7.11	6.3
	5.21	5.15	5.23	6.13	6.10	5.31	6.22		8.19	
	6.1	5.16	5.27	6.14	6.18	6.4	7.1		8.20	
	6.3	5.17	6.7	6.20	6.27	6.5				
		5.22	6.8	7.2	7.4					
		5.24	6.9	7.3						
		6.2	6.11							
		6.6	6.12							
		6.21	6.19							
		6.25	6.28							
Scale	11000	11800	12800	14187	13600	10000	16300	15600	21231	15000
Layer	11	12	13	14	15	16	17	18	19	20
Data	7.6	7.10	7.12	7.13	7.14	7.18	7.29	8.3	8.4	8.5
	7.7	7.19	7.16	7.27	7.15	7.30	8.14	8.7	8.17	8.10
	7.8	7.23	7.17	8.6	7.20	7.31	8.18	8.8		8.11
	7.9		7.22		7.21	8.1		8.9		
			7.24		7.28			8.13		
			7.25		8.16			8.15		
			8.2							
Scale	19000	20200	22000	25300	23500	22900	24200	26300	27700	28500

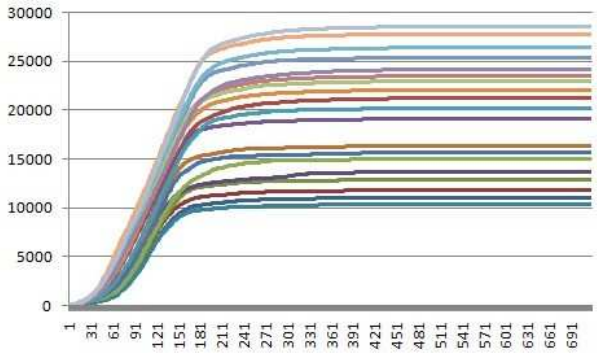

Fig. 3: Line graphs of discrete scales

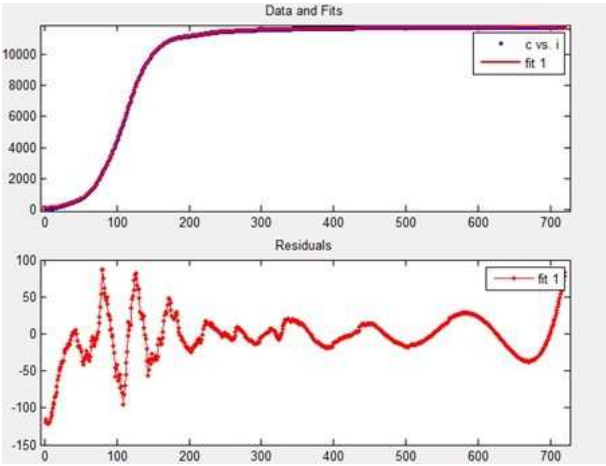

Fig. 4: Fitting and Residual curve

3.3 Gaussian fitting with three-phase

Compared with common fitting equations, the fitting accuracy of Gaussian fitting is the highest for the problem in this paper. Therefore, we made Gaussian fitting different data scales. And after large numbers of experiments, Gaussian fitting with eight polynomials has the minimum error among fitting experiments, so we choose Gaussian function with eight polynomials as the fitting function. Here the function is expressed as follows,

$$F(x) = A_1 * exp[-(\frac{x-B_1}{C_1})^2] + A_2 * exp[-(\frac{x-B_2}{C_2})^2] + A_3 * exp[-(\frac{x-B_3}{C_3})^2] + A_4 * exp[-(\frac{x-B_4}{C_4})^2] + A_5 * exp[-(\frac{x-B_5}{C_5})^2] + A_6 * exp[-(\frac{x-B_6}{C_6})^2] + A_7 * exp[-(\frac{x-B_7}{C_7})^2] + A_8 * exp[-(\frac{x-B_8}{C_8})^2]$$

Let's take the scale of 118000 for example and explain it in detail. Figure 4 shows fitting curve and residual curve. Obviously, the residual is a little big. We can consider a reasonable and scientific section way in order to reduce fitting residual.

In order to ensure the rationality of dividing stages, we take derivatives of tourist arrivals growth curve (See Figure 5). According to the change of its growth rate, we divide it into three stages and introduce the concept 'increment speed' of tourists to a scene. Taking derivatives of different scales of tourist arrivals growth, we find that the derivative values of all scales are the largest in 110th minute and become 0 in the 180th minute. In the 110th minute, 'increment speed' rises up to a maximum and in the 180th minute the 'increment speed' drops down to 0. So we take two 110 and 180 as two subsection point and divide it into three phases:(0,110), (110,180), (180,720).

Use Matlab to fit it and we can get corresponding Gaussian fitting equations of three phases. The parameters values of three stages' Gaussian equations are shown in table 2 below.

Table 2 Parameter values of Gaussian equation Coefficients (with 95% confidence bounds)

	First phase0-110	Second phase110-180	Third phase180-720
A1	667.3	155.4	1.17E+04
B1	114.3	181.3	787.2
C1	10.04	31.96	2056
A2	-19.35	-7.944	20.7
B2	105	163.2	227.9
C2	0.6545	5.696	7.678
A3	442.7	38.52	33.66
B3	99.09	157.4	243.6
C3	6.227	6.724	13.7
A4	201.7	21.59	21
B4	91.37	150.6	270.1
C4	5.366	2.93	10.44
A5	364.2	24.09	197.3
B5	83.11	147.5	247.7
C5	11.24	2.795	88.82
A6	-4.772E+06	-2519	11.37
B6	-4123	215	360.9
C6	1263	64.7	20.24
A7	1.15E+05	1358	132
B7	332.7	127.7	326.6
C7	126.3	28.58	159.6
A8	189.6	1.29E+04	265.3
B8	105.3	190.8	374.5
C8	3.899	82.77	245.4

Table 3 Goodness of fit

	First phase0-110	Second phase110-180	Third phase180-720
SSE	2.215e+004	4511	2699
R-square	0.9999	1	0.9998
Adjusted R-square	0.9999	1	0.9998
RMSE	16.24	10.13	2.291

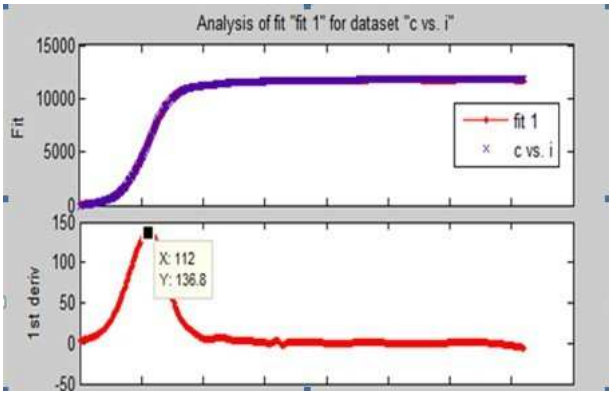

Fig. 5: Fitting and Residual curve

Table 3 illustrates that Gaussian fitting hold high fitting precision through the consistency check for this research.

Repeat these experiments and we can get each parameter value of Gaussian equations for all 20 discrete scales above. In order to cut down the paper, they are not listed here.

3.4 Weight calculating

The scales are more closer and the difference of curve graphics is more smaller, of course it's the same in reverse. In order to reduce the forecasting interference of too far-away scale to the experimental scale, we carry out a secondary clustering. 20 discrete scales are divided into four types and the result can be seen in table 4.

In order to examine the forecasting accuracy of the proposed model, this paper selected data of May 10, June 13, July 22, and August 8 respectively for scales of four types above to carry on the experiments.

From the historical data, we know that data scales of May 10, June 13, July 22 and August 8 are respectively 11044, 14301, 22126 and 26288 persons. Let us respectively mark the scale forecasting function of the four days as $F_1, F_2 F_3, F_4$. According to the proposed weight calculation steps based on improved OWA operator, we can get function expressions of $F_1, F_2 F_3, F_4$ as follows.

(1) $F_1 = 0.00116 f_1 + 0.66524 f_2 + 0.34359 f_3$, where f_1, f_2, f_3 represent Gaussian fitting functions of the scales of 10000, 11000 and 11800;

(2) $F_2 = 0.00021 f_4 + 0.31264 f_5 + 0.66492 f_6 + 0.00021 f_7 + 0,31264 f_8 + 0.0002 f_9$, where f_5, f_6, f_7, f_8, f_9

Table 3 Scales of four typest

Type	Scale
1	10000 11000 11800
2	12800 13600 14187 15000 15600 16300
3	19000 20200 21231 22000 22900 23500 24200
4	25300 26300 27700 28500

represent Gaussian fitting functions of the scales of 12800, 13600, 14187, 15000, 15600 and 16300;

(3) $F_3 = 0.00077f_{10} + 0.03932f_{11} + 0.18215f_{12} + 0.47983f_{13} + 0.21122f_{14} + 0.09270f_{15} + 0.02942f_{16}$, where f_{10}, f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16} represent Gaussian fitting functions of the scales of 19000, 20200, 21231, 22000, 22900, 23500 and 24200;

(4) $F_4 = 0.025356f_{17} + 0.62228f_{18} + 0.12414f_{19} + 0.00002f_{20}$, where f_{17}, f_{18}, f_{19}, f_{20} represent Gaussian fitting functions of the scales of 25300, 26300, 27700 and 28500.

3.5 Accuracy measures

No accuracy measure is generally applicable to all forecasting problems due to various forecasting objectives and data scales [22]. Let Y_t denote the observation at time t and F_t denote the forecast of Y_t .Then define the forecast error $e_t = Y_t - F_t$. In this paper, the following five measures of forecast accuracy are adopted to calculate the fitness of the model:

(1)Mean absolute error (MAE): Mean absolute error could reflect the forecast error more accurately, and overcome the defects of the average error in a certain extent. *MAE* is expressed as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|, t = 1, 2, \cdots, 720.$$

(2) Root mean square error $(RMSE)$: RMSE is popular and often chosen by practitioners because of its ease of use and its theoretical relevance in statistical modeling. *RMSE* is expressed as follows:

$$RMSE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}e^2}, t = 1, 2, \cdots, 720.$$

(3) Mean absolute percentage error $(MAPE)$: This criterion is less sensitive to large errors than *RMSE* and can be expressed as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{e_t}{Y_t}|, t = 1, 2, \cdots, 720.$$

(4) Mean square percentage error *MSPE*: *MSPE* is a little similar to *RMSE*, but *MSPE* is a percentage error. *MSPE* can be expressed as

$$MSPE = \frac{1}{n}\sum_{t=1}^{n}(\frac{e_t}{Y_t})^2, t = 1, 2, \cdots, 720.$$

4 Experimental results and analysis

This section presents experimental results of 4 experiments, including the four day of May 10, June 13, July 22, and August. Figure 6-13 show the experimental results. The forecasting result generated by the proposed model of May 10 is shown in Figure 6, where horizontal axis is time (minute) from 1 to 720, vertical axis is the number of tourist arrivals to the scenic entrance. The blue point denotes actual values of the total tourist arrivals number before each minute and the green point denotes the forecast values of tourist arrivals. Thus we can figure out the forecast value every minute. Figure 6 shows the forecast error of every minute on May 10.Results of the rest three days are similar to May 10(See Figure 8-13).

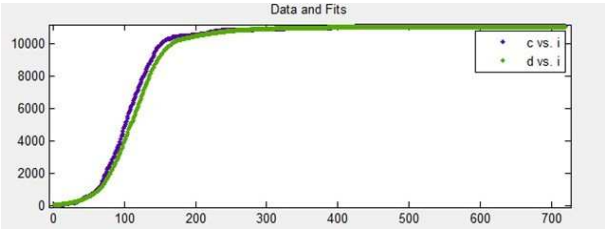

Fig. 6: Daily forecasting result generated by the proposed model (May 10)

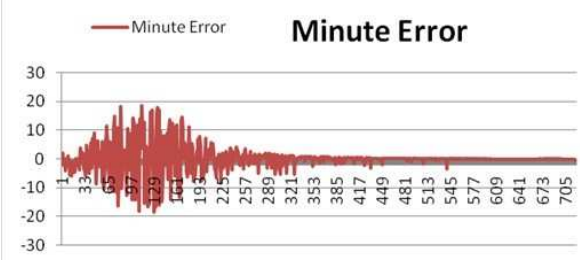

Fig. 7: 7 Minute Error (May 10)

From these experimental results above, the forecast error of every minute is presented clearly. And Table 4 shows the accuracy measure values of the four days by the proposed model.

In order to evaluate the proposed model further, we have chosen ARIMA model to compare with the proposed model

Table 4 Accuracy measures of four days (Proposed Model)

	MAE	RMSE	MSPE	MAPE
1	2.299148	0.167204	0.037198	0.510392
2	2.101395	0.155038	0.035977	0.428577
3	2.276626	0.154788	0.034848	0.475975
4	2.558617	0.172147	0.036526	0.422964

Table 5 Accuracy measures of four days (ARIMA)

	MAE	RMSE	MSPE	MAPE
1	5.629134	0.472337	0.126981	1.053046
2	6.060104	0.483465	0.068502	0.708167
3	7.191421	0.529687	0.07318	0.794814
4	7.206378	0.501461	0.100542	0.814325

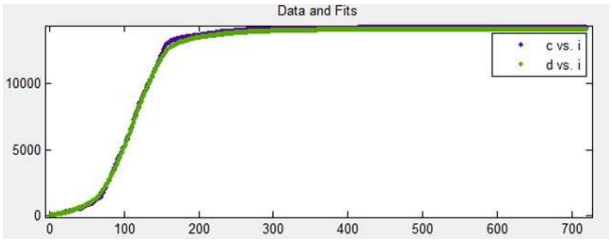

Fig. 8: Daily forecasting result generated by the proposed model (June 13)

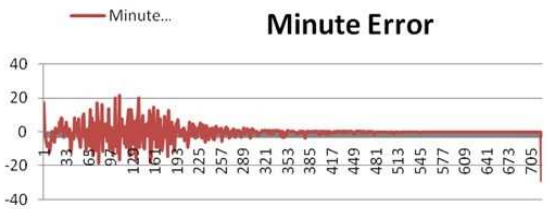

Fig. 9: Minute Error (June 13)

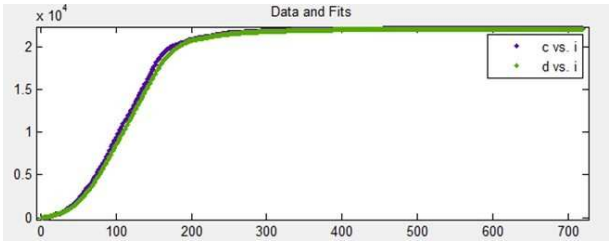

Fig. 10: Daily forecasting result generated by the proposed model (July 22)

(See Table 5). Comparing the experimental results and accuracy measures in Table 4 and Table 5, we can demonstrate that the performance of the proposed model is much superior to traditional ARIMA models.

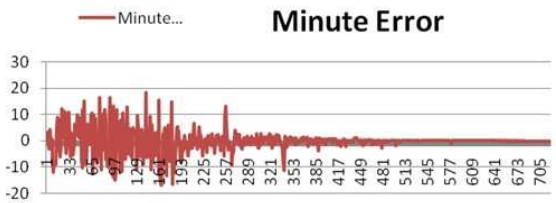

Fig. 11: 11 Minute Error (July 22)

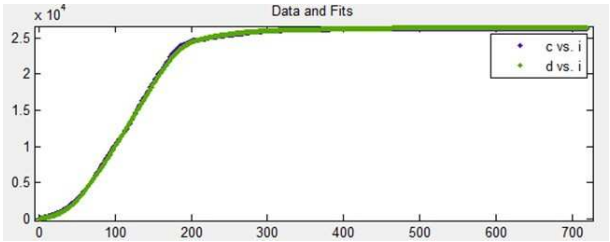

Fig. 12: Figure 7 Daily forecasting result generated by the proposed model (August 8)

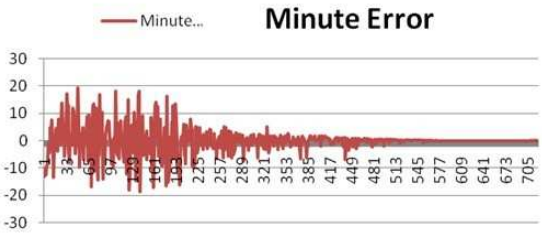

Fig. 13: Minute Error (August 8)

5 Conclusions

This paper put forward a new research question on forecasting tourist arrivals. Compared with precious studies, we forecast the tourist arrivals every minute of a day based on historical data. The dominant use of annual data, quarterly data

and monthly data mostly for previous analysis, is probably due to the fact that the explanatory variables at higher frequencies are not easy to obtain. But we obtained data benefiting from the use of RFID in scenes. And with the development of Information Technology such as RFID, GPS and so on, more and more scenes will put it to use. Therefore it is necessary and important that studying the problem of forecasting tourist arrivals at more refining time. This paper finds that different scales of tourist arrivals have a high similarity in arrival curves based on statistical regularities. Arrival curves show that the growth rate of tourists to the entrance increases as the time increases in the beginning stage. And after a period of growth period, the growth rate gradually slows down to zero, and finally total tourists approach to its certain limit under different scales.

Based on the cluster theory and stage forecasting model of tourism theory, this paper constructed a new forecast model of combing Gaussian fitting with three stages and weight solving to forecast real-time tourist arrivals. The new combined model and forecasting process are innovative. We took Jiuzhai Valley scenic area as an example to analysis and test the forecasting model .The model is proved valid with the empirical results and superior to traditional ARIMA models. However, this paper does not take residual er ror after curve fitting into consideration. In order to improve accuracy, we can consider using residual error to correct forecasting values in the future.

Acknowledgement

This work was supported by the Major International Joint Research Program of the National Natural Science Foundation of China (Grant no. 71020107027), the General Research Program of the National Natural Science Foundation of China (Grant no. 71371130), the National Natural Science Foundation of China (Grant no. 71001075).Humanities and Social Sciences project of The Ministry of education of China (Grant no.12YJC630023)

References

[1] George, A., Hyndman, R., Song H., International Journal of Forecasting, 27 (2011).
[2] World Tourism Organization, UNWTO World Tourism Barometer, 6 (2008).
[3] Haiyan Song, Gang Li, Stephen F. Witt, George Athanasopoulos,) International Journal of Forecasting, 27 (2011).
[4] Box, G., Jenkins, G. and Reinsel, G., New Jersey: Wiley, (2008).
[5] Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D., Journal of Forecasting, 24 (2005).
[6] Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S., International Journal of Forecasting, 18 (2002).
[7] Ord, J. K., Koehler, A. B. and Snyder, R. D., Journal of the American Statistical Association, 92 (1997).
[8] Assimakopoulos, V. and Nikolopoulos, K., International Journal of Forecasting, 16 (2000).
[9] Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D., Berlin, Heidelberg: Springer-Verlag, (2008).
[10] Rob Law, Tourism Management, 20 (1999).
[11] Cho, V.,Tourism Management, 24 (2003).
[12] Law, R. Tourism Management, 21 (2000).
[13] Tugba, T. T., Casey, M. C., Neural Networks, 18 (2005).
[14] Palmer, A., Montano, J. J., Sese, A. , Tourism Management, 27 (2006).
[15] Au, N., Law, R., Journal of Travel Research, 39 (2000).
[16] Au, N., Law, R., Annals of Tourism Research, 29 (2002).
[17] Song, H., Li, G. , Tourism Management, 29 (2008).
[18] Li, G., Song, H., Witt, S. F., Journal of Travel Research, 44 (2005).
[19] Matteo Palmucci, Mario Giordano, Environmental and Experimental Botany, 75 (2012).
[20] Yu Wei, Mu-Chen Chen, Transportation Research Part C, 21 (2012).
[21] Chen, H., LIiu, C., Forecasting, 22 (2003).
[22] De Gooijer, J., Hyndman, R., International Journal of Forecasting, 22 (2006) .

Lifei Yao is a Graduate student in Sichuan University for majoring in management science and engineering, and a member of Information and Business Management Institute of Sichuan University. Her research interests are prediction, evaluation, decision control and vehicle scheduling.

Ruimin Ma is a doctoral student in Sichuan University for majoring in management science and engineering, and a member of Information and Business Management Institute of Sichuan University. His researches mainly relate to simulation, multi-objective decision, and vehicle scheduling, etc.

Maozhu Jin is an instructor of Business School,the tutor of MBA operations management and innovation and entrepreneurship management in Sichuan University. He has been engaged in the teaching of core curriculums such as operations management and management consulting. His current research interests include the areas of operations management, organizational process reengineering, strategic management, service operations management, platform-based mass customization and risk management. He has published two books and over ten research papers in authoritative journals of high quality both at home and abroad, and ten of them are retrieved by SCI and EI.

Peng Ge is an instructor of Business School,the tutor of MBA operations management and innovation and entrepreneurship management in Sichuan University. He has been engaged in the teaching of core curriculums such as operations management and management consulting. His current research interests include scientific management, industrial engineering, tourism management and so on. Over ten research papers in authoritative journals of high quality have been published at home and abroad, and many of them are retrieved by SCI and EI.

Peiyu Ren is a Professor, PhD Supervisor, currently acting as the Director of Information and Business Management Research Institute of Sichuan University. He has presided over and completed five surface projects of National Natural Science Foundation of China, being in charge of project research of Projects 863, 985 and 211, having published 15 books, monographs and more than 100 academic papers, including SCI, EI and CSSCI. Contact him at renpy.scu163.com.