Research on Prediction of Visitors Quantity Based on GM(1, 1) Method

Cai-Hong SUN

School of Business

Huanggang Normal University, HNU

Huanggang, China

E-mail: 351851979@qq.com

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 visitor's quantity according to relevant influencing factors is essential to the development of tourism economy. This paper firstly introduces several common prediction method principle, then according to the characteristics of visitors quantity, builds GM (1, 1) prediction model. The result shows that the method is feasible and effective, the results is accurate.

Keywords- visitors quantity; prediction; GM(1, 1) method

I. INTRODUCTION

Forecasting is the foundation and basis of decision. Use appropriate prediction to get tourism economy development rule and provide scientific basis for decision-making in the course of tourism economic development. Avoid economic environmental loss caused by decision-making errors, it is necessary to realize the sustainable development of tourism. Especially 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.

As everyone knows, tourism science is a young discipline only 20 years of history in China, 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 of lag of 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 "atrial information known, part information unknown" "small sample, poor information, uncertain" system research. Based on this, this paper tries to

use the grey system theory to predict visitor's quantity, and take Huanggang city as an example to demonstrate the feasibility of method.

II. ESTABLISHMENT OF GREY PREDICTION GM (1,1) MODEL

A. Modeling of 1. GM (1,1) Model

Set time series $X^{(0)}$ have an observed value $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n),\}$

Generate a new sequence by accumulating $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n),\}$ among this,

$X^{(1)}(k) = \sum_{i=1}^k X^{(0)}(i), k = 1, 2, \dots, n$ the differential equations for the linear dynamic grey prediction model GM (1, 1) of a single sequence is $\frac{dX^{(1)}}{dt} + \alpha X^{(1)} = \mu$.

Among them a is the development of grey number, reflecting the size of the development of the system, μ is endogenous control gray number. GM (1, 1) model applicable conditions as shown in Table I.

TABLE I. GM(1, 1) MODEL APPLICABLE CONDITIONS

Range	Remarks
$-\alpha < 0.3$	Medium and long term forecast
$0.3 < -\alpha < 0.5$	Short term prediction
$0.5 < -\alpha < 0.8$	With the short-term forecast
$0.8 < -\alpha < 1$	Residual error correction model

Set $\hat{\alpha}$ is the parameter vector to be estimated, $\hat{\alpha} = \begin{pmatrix} \alpha \\ \mu \end{pmatrix}$

least square method can be used to solve. Solution: $\hat{\alpha} = (B^T B)^{-1} B^T Y_n$. Solving differential equations, you can have to predict the model

$$\hat{X}^{(1)}(k+1) = \left[X^{(0)}(1) - \frac{\mu}{\alpha} \right] e^{-\alpha k} + \frac{\mu}{\alpha}, \quad k=0,1,2,\dots,n.$$

B. Model Checking

In order to guarantee the credibility of the model, we have residual test, correlation test and posterior difference test of grey prediction.

1) Residual Test

Set $\Delta^{(0)}(i) = |x^{(0)}(i) - \hat{x}^{(0)}(i)|$ is absolute error sequence for the original series and the forecast value, Relative error sequence is $\varnothing(i) = \frac{\Delta^{(0)}(i)}{x^{(0)}(i)} * 100\%$ Fitting accuracy is

$P_0 = 1 - \frac{1}{n} \sum_{i=1}^n \phi(i)$, If $\varnothing(i)$ can control in less than 8%, it may be considered by the model residual test.

2) Correlation Test

Set

$\hat{X}^{(0)}(k) = \{\hat{X}^{(0)}(1), \hat{X}^{(0)}(2), \dots, \hat{X}^{(0)}(n)\}$, $\hat{X}^{(1)}(k) = \{\hat{X}^{(1)}(1), \hat{X}^{(1)}(2), \dots, \hat{X}^{(1)}(n)\}$, then correlation coefficient is defined as

$$\eta(k) = \frac{\min \min |\hat{X}^{(0)}(k) - X^{(0)}(k)| + \rho \max \max |\hat{X}^{(0)}(k) - X^{(0)}(k)|}{|\hat{X}^{(0)}(k) - X^{(0)}(k)| + \rho \max \max |\hat{X}^{(0)}(k) - X^{(0)}(k)|}$$

Among them, ρ is the resolution, generally take 0.5,

$r = \frac{1}{n} \sum_{k=1}^n \eta(k)$ is known as the correlation of $X^{(0)}(k)$ with $\hat{X}^{(0)}(k)$.

3) Testing for Posterior

Set S_1 is standard deviation of the original sequence, S_2 is standard deviation sequence of absolute error, Variance ratio is $C = S_2/S_1$,

Small probability of error is $P = p\left\{|\Delta^{(0)}(i) - \overline{\Delta^{(0)}}| < 0.6745 \times S_1\right\}$, set $e_i = |\Delta^{(0)}(i) - \overline{\Delta^{(0)}}|$, $S_0 = 0.6745 \times S_1$, then $P = p\{e_i < S_0\}$, Criteria as shown in Table II

TABLE II. POSTERIOR VARIANCE TEST CRITERIA

P	C	Effect
>0.95	<0.35	Good
>0.80	<0.50	qualified
>0.70	<0.65	Barely qualified
≤0.70	≥0.65	Unqualified

If the residual test, Correlation test, posterior variance test can be passed, can be predicted by the model, otherwise, we need to be corrected residuals.

III. TRAFFIC FORECAST THE NEXT FIVE YEARS

A. Segment Traffic Raw Data

Due to the volume of business segments A raw data as time-series data too, are shown in Table III, are not suitable for use ARIMA forecasting model; and gray prediction GM (1, 1) is adapted to predict the small sample size case.

TABLE III. DIVISION OF THE PAST FOUR YEARS THE VOLUME OF BUSINESS

	Three years ago	Two years ago	Previous year	Currently
volume of business	216411	283499	382723	535430

B. Forecast Result

We use Matlab6.5 prepared a GM (1, 1) prediction program, after the call, the operation results are as follows,

1) gray number of development $\alpha = 0.0$, Endogenous control variables $\mu = 166710$, Chart 6-1 shows that the conditions applicable model, the development of factor $\alpha = 0.0 < 0.3$, the model is suitable to do long term forecasting.

2) Forecast result is as shown in Figure I

Figure 1. A division of the past four years the volume of business

3) Model Checking

(1) Residual test. By calculation, the fitting precision of about $P_0 = 98.96\%$, considered by residual test.

(2) Correlation test. By calculating the degree of relevance to 0.6648 > 0.6, considered by Correlation test.

(3) Posterior variance test. After calculation,

$C = 0.0338 < 0.35$; but $e_i = [4167.2 \ 1686.7 \ 3380.4$

$5860.8]$, $S_0 = 93620$, so $P = p\{e_i < S_0\} = 1 > 0.95$, passed

Posterior variance test.

4) A division next five years, traffic forecast value.

By Grey Prediction GM (1, 1) model and Matlab software, we predicted the next five years, A Division of traffic predicted values shown in Table IV.

TABLE IV. A DIVISION NEXT FIVE YEARS, THE TRAFFIC FORECAST

	Future first year	In the next second	Future third year	Future fourth year	Future fifth year
volume of business	722898	994202	1367791	1881580	2588215

Through Table IV, combined with C and P values, our analysis: the model has good prediction accuracy and practical value, predicted results can be used.

IV. CONCLUSION

Accurate prediction of visitor's 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 can achieve satisfactory results, this is the learning direction in the study of combined prediction method to the direction of learning in the future.

REFERENCES

- [1] K•C•Luk, J•E•Bal, I A•Sharma•2000, A Study of Optimal modelling and Spatial Inputs to Artificial neural network for Rainfall Forecas-ting, Journal of Hydrology 227: 56-65.
- [2] G-B. Zhao, Y-F Shi, W-F Duan, "Com-puting fractal dimension and the Kolmogorov entropy from chaotic time series," Chinese Journal of Computation Physics, vol. 16, Mar.1999:pp. 309-315.
- [3] J-H. Lü, J-A. Lu, S-H. Chen, Chaotic Time Series Analysis and its Applications, Wuhan: Wuhan University Press, 2002.
- [4] K. Chen, B-T Han, "A survey of state space re-construction of chaotic time series analysis," Computer Science, vol. 32, Apr. 2005: pp67-70.
- [5] X-Y Wen, M-H Niu, A neural network based nonlinear combination forecasting method, System engineering theory and Practice, Dec.1994.
- [6] Q Zhang, Study on the Neural network based optimal combination forecasting method, System engineering theory and Practice, 2001(9).
- [7] T. Luo, "Optimization Grey Model", Chinese Academy of Engineering, Aug.2003.
- [8] K. Chen, B-T Han, "A survey of state space re-construction of chaotic time series analysis," Computer Science, vol. 32, Apr. 2005: pp67-70.
- [9] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics delay times and embedding windows," Physical D Nonlinear Phenomena, vol. 127, Feb. 1999: pp. 48-60.