IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 5, May 2015.
www.ijiset.com
ISSN 2348 – 7968

The Time Series Forecast Model Research and Disseminates in The Question Application at Disease

Zhang Di,Wu Wenrui, Ha Shuang, Yang Aimin*, Han Yang, Li Tan

North China University of Science and Technology, Tangshan, 063000, China

Email: aimin_heut@163.com

Abstract

We have focused on the time series forecasting model, predict the next period of time the number of features, according to the result which will obtain, the union residual error analytical control, verify the accuracy of predictions, achieve the valid forecast to the viral spread condition the goal. Take Africa's Guinean country as an example, carries on the data simulation analytical control, obtain this national infection population and the casualty forecast value and the actual value are close, carry on the residual error analytical control, the nearly all residual error confidence interval all contains 0, prove the accuracy of forecast value , has confirmed the time series model rationality well. According to forecast population which above obtains, calculate the need of the medicine in this local period of time, provide the reliable data basis for the formula.

Keywords:*Time Series Forecast,Residual Error Analytical Control,Confidence Interval.*

1. Introduction

Disease dissemination has become in the modern people heart the nightmare, has the active control disease dissemination also become how one of topics which the modern people most cared about, the domestic and foreign scholars have done the massive research in this aspect, for example, Li Wei, Bi Guihong, Zhang Shouming[1]proposes in 2012 disseminates the model based on Agent dynamic network disease, Mainly uses in studying disease to disseminate the proliferation dynamic process in the crowd; Song Tong and Li han in 2011 proposed one kind of new chaos time series forecast model[2-3]wavelet echo condition network, this model may overcome in effectively the traditional echo condition network model the universal existence morbid state matrix question, increased the chaos time series forecast precision; Jaeger and Haas proposed in 2004 one kind of new time series model[4], namely the echo condition network, it causes the network through the reserve pond network interior sparse connection to have the memory ability; George E.P.Box and Gwilym M.Jenkins et al. have published the monograph called "Times series Analysis Forecasting and Control"[5]in 1970, from now on symbolized the time series forecast analysis, including the data processing, the model establishment, the model parameter estimation as well as the forecast and the control related one whole set theory already took shape and gradually improves; In order to solve the non-steady time series forecast problem, George E.P.Box and Gwilym M.Jenkins proposed from the return conformity running mean model; Yuan Zhenzhou proposed using the ARIMA model forecast 1989-1994 year Beijing Railroad Bureau's coal freight transportation total quantity data, the forecast effect is good; In 1987, Granger and Engle first proposed cooperates the reorganization theory, this theory goal lies in balances two or the many non-steady sequences, and cooperates the entire relational with the existence the time series establishment error correction model[6].

Above literature indicated that, the use time series forecast model analysis data, estimated methods and so on parameter are already studied by the multi-people, but unifies the use the time series and the variance analysis to report rarely seen actually, therefore, the author thought had the necessity to study the time series forecast model with emphasis, understood it in the forecast future in period of time the population change characteristic, according to the result which will obtain, the union residual error analytical control, the confirmation forecasts the result the accuracy.

2. Infection Population Forecast Model

2.1 Model preparation

The time series are according to the time arranged in order, along with the time variation also the interdependence data sequence. An analysis time sequence method constitution data analysis important domain, namely time series analysis. If in the forecast time scope, does not have changes also the stochastic change called σ^2 variance suddenly is small, and has the reason to think and the present evolution tendency will continue when will develop the future, the available some rule of thumb carries on the forecast. In this article period of revolution sequence forecast index smoothing procedures, in which index smoothing procedures basis smooth number of times difference, also divides into time index smoothing procedures, two index smoothing procedures and three index smoothing procedures and so on.

Although index smoothing procedures have overcome the moving average method shortcoming. But when the time

ISSN 2348 – 7968

series change has the straight line tendency, carries on the forecast with index smoothing procedures, still had the obvious lag deviation. Therefore, also must perform to revise. The revision method and the tendency moving average method is same, namely makes two indices to be smooth again, use lag deviation rule establishment straight line tendency model. This is two index smoothing procedures. Its formula is:

$$\begin{aligned} S_t^{(1)} &= \alpha y_t + (1-\alpha) S_{t-1}^{(1)} \\ S_t^{(2)} &= \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)} \end{aligned} \quad (1)$$

In the formula $S_t^{(1)}$ is an index smooth value; $S_t^{(2)}$ is two index smooth values. When time series $\{y_t\}$, starts from some time has the straight line tendency, the similar tendency moving average method, the available straight line tendency model carries on the forecast.

$$\hat{y}_{t+T} = a_t + b_t T, T = 1,2,\cdots \quad (2)$$

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) \end{cases}$$

When time series change performance for conic section tendency, then needs to use three index smoothing procedures. Three indices smooth are in two index smooth foundations, again carries on one time smooth, its formula is:

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha) S_{t-1}^{(3)} \end{cases} \quad (3)$$

In the formula $S_t^{(3)}$ is three index smooth values.
The prediction model of the cubic exponential smoothing method is:

$$\hat{y}_{t+T} = a_t + b_t + c_t T^2, T = 1,2,\cdots \quad (4)$$

In which

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^2 - S_t^{(3)} \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(3)}\right] \\ c_t = \dfrac{\alpha^2}{2(1-\alpha)^2}\left[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right] \end{cases} \quad (5)$$

2.2 The Determination of the Number of Infections

According to the Guinean country about the Egyptian Borah virus epidemic situation short-term data, will obtain the change chart which as shown in Fig. 1.

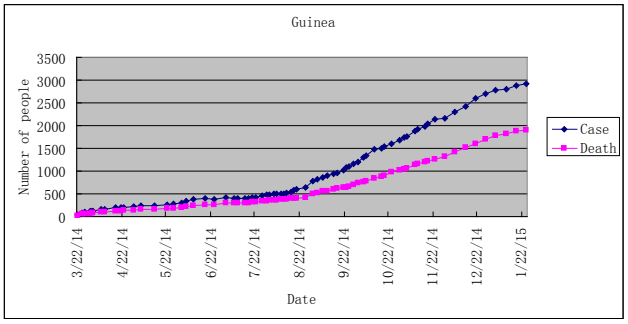

Fig.1 This national recently data statistics line chart

Take Guinea as an example, the situation of Ebola outbreak will be updated once a week, if the data are screened once every 7days, we can get as shown in table 1:

Tab. 1 The Data of Ebola outbreak in Guinea

Release date	3/22/14	3/25/14	4/1/14	4/10/14	4/17/14	4/23/14
Case	49	86	122	157	197	208
Death	29	59	80	101	122	136
Release date	4/30/14	5/5/14	5/14/14	5/23/14	6/2/14	6/10/14
Case	221	231	233	258	291	372
Death	146	155	157	174	193	236
Release date	6/18/14	6/24/14	7/2/14	7/8/14	7/15/14	7/23/14
Case	398	390	413	408	406	427
Death	264	270	303	307	304	319
Release date	7/30/14	8/6/14	8/13/14	8/20/14	8/26/14	9/3/14
Case	472	495	519	607	648	823
Death	346	367	380	406	430	522
Release date	9/10/14	9/17/14	9/23/14	10/1/14	10/7/14	10/17/14
Case	899	965	1074	1199	1350	1501
Death	568	623	648	739	778	886
Release date	10/24/14	10/30/14	11/4/14	11/11/14	11/18/14	11/23/14
Case	1598	1675	1760	1919	2047	2134
Death	981	1022	1054	1166	1214	1260
Release date	12/7/14	12/14/14	12/21/14	12/28/14	1/4/15	1/11/15
Case	2292	2416	2597	2707	2775	2806
Death	1428	1525	1607	1708	1781	1814

Note: Date format, mm/dd/yy

According to the prediction model of time sequence in 5.1.1, we can use SPSS to fit the data in table 1 into linear Function and quadratic function, then we can get the both R-square values is 0.901, it shows the fitting degree is comparatively high, what' s more, we can also find the six curves all tend to be the quadratic function, then we can say cubic exponential smoothing method is better in prediction.

Assuming $\alpha = 0.3$, The initial value is $S_1^{(0)} = S_2^{(0)} = s_3^{(0)} = \dfrac{y_1 + y_2 + y_3}{3} = 85.67$.

ISSN 2348 – 7968

Computing $S_t^{(1)}$, $S_t^{(2)}$, $S_t^{(3)}$ using MATLAB, as shown in Table 2. Where to get:

$a_{45} = 2893.42$, $b_{45} = 77.87$, $c_{45} = -1.76$

Thus, gain the model when t is 45:

$$\hat{y}_{45+T} = 2893.42 + 77.87T - 1.76T^2$$

Tab.2 The number of patient in Guinea forecasted through cubic exponential smoothing method

Release date	t	Case y_t	a_t	b_t	c_t	Estimates of the y_{t+1}
3/22/14	1	49	85.67	0.00	0.00	85.67
3/25/14	2	86	61.58	-8.41	-0.50	52.67
4/1/14	3	122	74.57	-1.75	-0.05	72.77
4/10/14	4	157	105.11	9.45	0.62	115.19
4/17/14	5	197	142.66	20.29	1.18	164.13
4/23/14	6	208	185.73	30.20	1.63	217.56
…	…		…		…	
1/11/15	43	2806	2794.20	117.99	0.56	2912.76
1/18/15	44	2871	2842.62	94.62	-0.88	2936.36
1/25/15	45	2917	2893.42	77.87	-1.76	2969.53

We can get the actual case value and predictive case value as shown in Fig. 2

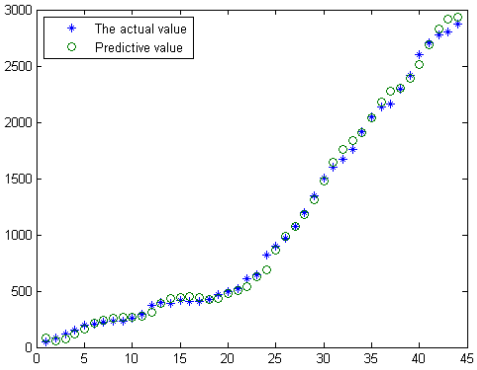

Fig.2 The comparison between actual case value and predictive case value in Guinea

When t is 43,the model is:

$$\hat{y}_{43+T} = 2794.20 + 117.99T + 0.56T^2$$

Then we can predict, in Guinea, 1/18/15 case is $\hat{y}_{44} = 2936$, it is almost the same as the actual value 2871.;1/25/15 case is $\hat{y}_{45} = 2969$, it is almost the same as the actual value 2917.

Repeat the same step:

We can predict, in Guinea, 1/18/15 deaths is 1917, it is quite close to the actual value 1876; we can predict, in Guinea, 1/25/ 15 deaths is 1961, close to the actual value 1910.

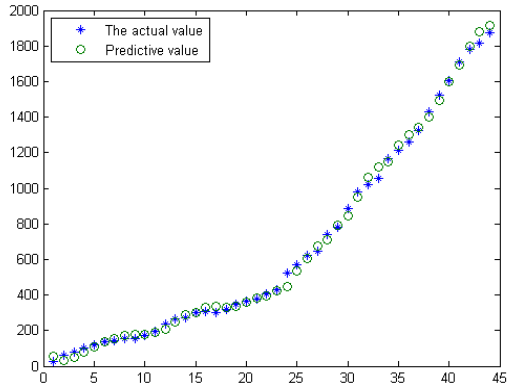

Fig.3 The comparison of between actual death value and predictive death value in Guinea

According to the case value and death value, we can get the number of people who need treatment is 1008, the number is quite close the actual value 1007. So this method can be used to predict the number of people who need treatment.

2.3 The Error of The Final Examines

This article according to Table 2 the actual infection population and the corresponding time forecast infection population value, unifies the Matlab programming to obtain the forecast value and actual value residual error chart 4 as follows:

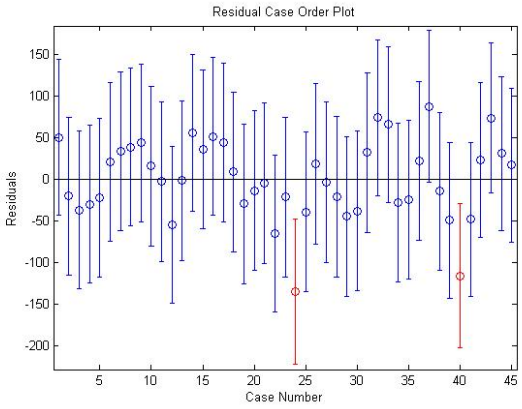

Fig. 4 The actual values and forecast value residual error charts

Obtains each time residual error value and residual error confidence interval as shown in Table 3 with Matlab:

Table 3 each time residual error values and residual error confidence interval

ISSN 2348 – 7968

Residual error value	Confidence interval	
	Lower limit	Upper limit
-36.67	-43.0046	144.6823
33.33	-114.8615	75.2715
49.23	-130.9979	58.3744
41.81	-124.4935	65.5109
32.87	-116.547	74.0738
-9.56	-74.3358	116.3458
...	...	...
-106.76	-16.3683	163.7644
-65.36	-60.9539	123.323
-52.53	-74.7057	109.8387

May see by on table, the nearly all residual error confidence interval all contains 0, explained the time series model which can conform to the primary data well, achieves the accurate forecast the effect.

3. Radiation Model of Predicting Disease Spread

The world medical association has announced that their new medication could stop Ebola and cure patients whose disease is not advanced. In order to Estimate the amount of drug needed, we can divided the infected into three groups: the early, middle and advanced patients, the patients in incubation period can′ t use drug for treatment, and also couldn′ t use vaccine to prevent through injection.

First, assume that each person's ability to fight the virus is the same, and each people have the same infection rate. Then the amount of drug that individual patient need varies according to the degree of infection. So, suppose the amount of drug that the early patients need is a, what the middle patients need is B(the amount has nothing to do with the age, only connected with the degree of infection) , what the advanced patients need is C (suppose all the patients will buy the same drug.)

Second, since the time ratio of the early, middle, advanced period remains the same, we suppose the amount of the patients in the three periods is separately: x , y , z , the ratio is $\alpha+\beta+\gamma=1$. What′ s more, suppose the total amount of drug is n_m , the total amount of patients is n_p Accordingly, the following equations can be obtained

$$\begin{cases} n_p = \dfrac{x}{\alpha} = \dfrac{y}{\beta} = \dfrac{z}{\gamma} \\ \alpha+\beta+\gamma=1 \\ n_m = ax+by+cz \end{cases} \tag{6}$$

The above equations simplify:

$$n_m = (a\alpha+b\beta+c\gamma)n_p = [(a-c)\alpha+(b-c)\beta+c]n_p \tag{7}$$

Data Simulation:

Hypothesis $a=15mL$, $b=25mL$, $c=20mL$, $\alpha:\beta:\gamma=2:5:3$ .In Guinea, for example, to predict the total number of countries is $n_p=1009$ in need of treatment in 2/1/15.It can be predicted that country 2/1/15 requires the total amount of drug is $n_m=21.694L$.

4. Conclusions

This article gives an example the return model needs the independent variable data quite is deficient, but must forecast the variable the historical data quite is complete, reflected sufficiently its change tendency the situation, conforms to the time series forecast model, also the time series forecast model compares with the return forecast model, this method does not need to spend the energy to determine between the quotation variable the causal relation, but will only have the historical tendency which determined the time series model outward to extend then may forecast the future the change, simpler. And this article unified the residual error analysis to carry on the analytical control to the data, has proven the forecast data reliability fully.

Acknowledgments

Work is supported by College Students' Innovative Project in North China University of Science and Technology(X2 015262).

References

[1] LI Wei,BI Gui-hong, ZHANG Shou-ming, "Epidemic Propagation Model Based on Agent Dynamical Network",Computer Engineering,Vol.38 ,No. 1 ,2012,pp.27-36.

[2] Song Tong, Li Han, "Chaotic time series prediction based on wavelet echo state network",Acta Phys,Vol.61,No.8,2012, pp.139-151.

[3] Zhifeng YANG, Yang ZHAO, Guobing FAN, "A New Iteration Algorithm for Solving the Diffusion Problem In non-Differentiable Heat Transfer",Thermal Science, Vol.19,No.suppl.1,2015, pp.105-108.

[4] Song A,Lu J, "Evolving Gaussian RBF network for nonlinear time series modelling and prediction",Electronics Letters, Vol.34,No.21,1998,pp.1241-1243.

[5] Shu XU, Xiang LING, Yang ZHAO, Hassan Kamil JASSIM, "A Novel Schedule for Solving the Two-Dimensional Diffusion Problem in Fractal Heat Transfer",Thermal Science, Vol.19,No.suppl.1,2015, pp.99-103.

[6] Jarosawski, Janusz, "Improvement of the Umkehr ozone profile by the neural network method". International Journal of Remote Sensing, Vol.34, No.15, 2013, pp.5541-5550.

ISSN 2348 – 7968

Zhang Di He is studying in North China of Science and Technology. His major is Safety Engineering.

Wu Wenrui He is studying in North China of Science and Technology. His major is Safety Engineering.

Ha Shuang She is studying in North China of Science and Technology. Her major is Computer Science and Technology.

Han Yang He is studying a master's degree in North China of Science and Technology. His major is Mathematics Specialty.

Li Tan He is studying in North China University of Science and Technology. His major is Surveying and mapping engineering.