ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 153

International Conference on Applied Mathematics, Modeling and Simulation (AMMS 2017)

A Set of Time Series Forecasting Model Based on the Difference

Hao Feng[1], Hongxu Wang[2], Chengguo Yin[3], Xiaoli Lu[2] and Xiaofang Fu[3]

[1]College of Marine Science and Technology, Hainan Tropical Ocean University, Sanya, China

[2]Commercial College, Hainan Tropical Ocean University, Sanya, China

[3]College of Computer Science and Engineering, Hainan Tropical Ocean University, Sanya, China

***Abstract*—A set of time series forecasting models based on difference is proposed (SD). For a time series, it can select the best time series forecasting model in SD by using the automatic optimal search method. For example, when forecast enrollments data of University of Alabama in 1971~1992, it can select the best time series forecasting model Dq(0.000003,0.000003) in SD by using the automatic optimal search method, and can gain the MSE=0 and AFER=0%. The fact that the prediction accuracy of the existing fuzzy time series prediction model is not very high has been fundamentally improved.**

Keywords-the difference; the function Xq(s,t) of SD's sum of fraction; the inverse function Zq(s,t) of SD's sum of fraction; the prediction function Dq(s,t) of SD; time series

I. INTRODUCTION

Mining the inherent rule of time series data is one of the tasks of studying time series [1]. For the regular time series, the model and method of analysis of classical time series have already solved many reality problems [1]. For the rule-less time series, the research history may take longer. In the early 1927, Yule [2] analyzed the time series of Wolfer 's sunspots, and draw an important conclusion about sunspot number will appear a maximum value in every 9.5 years to 11 years. In 1993, Song and Chissom [3,4] apply the fuzzy set theory[5]. Taking the lead in studying fuzzy time series prediction model, and successfully simulated the prediction of University of Alabama 1971~1992 enrollment, but the prediction accuracy is not too high. We noted that the sunspot number series and the registration number series are time series, such as the registration number [4], as shown in table (2), and its distribution map show as follow.

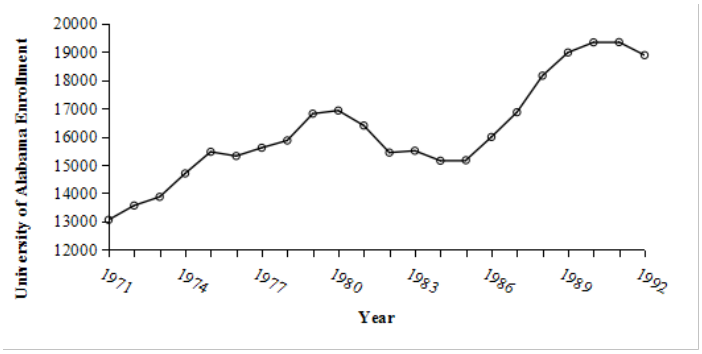

FIGURE I. THE DISTRIBUTION MAP OF THE ENROLLMENT OF UNIVERSITY OF ALABAMA 1971~1992

Showing in figure 1. Actually, it's a choppy, chaotic sequence of times series. These two cases are rule-less time series. So far, the fuzzy time series forecasting model has been put forward (the prediction model mentioned in document [6-15] is a part of it). In 2009 Stevenson and Porter[6] proposed the forecasting model of Fuzzy time series , when simulate and predict the enrollment number of University of Alabama, get the average prediction error and mean square error AFER=0.57% MSE=21575[6], although it have been the best prediction accuracy since 2009, but still not too high. However, the concept of fuzziness used in this paper is valuable. In this paper, a set of time series forecasting models based on difference (SD) is proposed. And it put forward the automatic searching method and the best time series forecasting model. When forecast and simulate the enrollment number of University of Alabama 1971~1992 years, it can search the best SD time series forecasting model $D_q(0.000003,0.000003)$ by using the automatic searching method.

II. A SET OF TIME SERIES FORECASTING MODEL BY SD

Definition 1: Set the historical data discourse domain of time series as $V=\{V_1,V_2,\ldots,V_n\}$.The Calculation formula about the historical data of difference as $W_q=V_q-V_{q-1}$, Use the formula to calculate and gain the difference discourse domain of historical data $W=\{W_2,W_3,\ldots,W_n\}$.

Definition 2: Set the historical data discourse domain of time series as $V=\{V_1,V_2,\ldots,V_n\}$.The historical data of difference as $W=\{W_2,W_3,\ldots,W_n\}$ .For every certain $q\in\{2,3,\cdots,n\}$, it is in the definition of W.

$$X_q(s,t)=\begin{cases}\dfrac{1}{W_2}+\dfrac{1}{W_3}, q=2\\ \dfrac{s}{W_{q-1}}+\dfrac{1}{W_q}+\dfrac{t}{W_{q+1}}, 3\le q\le n-1,\\ \dfrac{s}{W_{n-1}}+\dfrac{1}{W_n}, q=n\end{cases}$$

$$Z_q(s,t)=\begin{cases}\dfrac{1+t}{\dfrac{1}{W_2}+\dfrac{t}{W_3}}, q=2\\ \dfrac{s+1+t}{\dfrac{s}{W_{q-1}}+\dfrac{1}{W_q}+\dfrac{t}{W_{q+1}}}, 3\le q\le n-1\\ \dfrac{s+1}{\dfrac{s}{W_{n-1}}+\dfrac{1}{W_n}}, q=n\end{cases}$$

$$D_q(s,t)=V_{q-1}+Z_q(s,t).$$

Where independent variable $s\in(0,1)$ and $t\in(0,1)$, $Z_q(s,t)$ is the corresponding inverse function about the function $X_q(s,t)$ of SD's Sum of fraction, thus $D_q(s,t)$ is the forecasting function of SD.

When the independent variable s and t are select the specific values in their discourse domain, can build up a forecasting formula $D_q(s,t)$ which is a time series forecasting model. For a time series, can apply the forecasting model $D_q(s,t)$ to simulate the historical data of forecasting time series.

Definition 3: Set the historical data discourse domain of time series as $V=\{V_1,V_2,\ldots,V_n\}$.The historical data of difference as $W=\{W_2,W_3,\ldots,W_n\}$.When the independent variable s and t are select the specific values in their discourse domain, can get endless time series forecasting models $D_q(s,t)$. We call the whole time series forecasting model $D_q(s,t)$ as A Set of Time Series Forecasting Model Based on the Difference. Simplify it as SD, and the general element of SD is $D_q(s,t)$.

Definition 4: The time series forecasting model $D_q(s,t)$ is call the best, if apply this forecasting model to simulate and forecast the historical data of some time series, it can get the MSE=0 and meanwhile AFER=0%.

In which MSE(Mean Square Error) is $MSE=\frac{1}{n-1}\sum_{q=2}^{n}(V_q-D_q)^2$, AFER(Average Forecasting Error Rate) is $AFER=\left(\frac{1}{n-1}\sum_{q=2}^{n}|V_q-D_q|/V_q\right)\times 100\%$.

Definition 5: For a time series, selecting the best time series forecasting model in SD by using computer automatic search method. We call this method as Automatic optimization search method. The steps for implementation are show as follow. For a time series, From a decimal number as the starting point of programming, programming, search, calculation,... ,Until you filter out the best time series prediction model $D_q(s,t)$ in SD.

Example1: The implementation case of automatic optimization search method. If simulate and forecast the enrollment data of University of Alabama 1971~1992, we choose $s=t=0.0003$ as the beginning point and follow to choose $s=t=0.0003,0.00003,0.000003,\cdots$ □by programming, search, calculation,...,Until you filter out the best time series prediction model $D_q(s,t)$ in SD. Firstly, we get the table 1, but the AFER $\neq$ 0% and MSE $\neq$ 0, and then we continue to search ,calculate and get table 2, but AFER $\neq$ 0% and MSE $\neq$ 0. We go on to search, calculate and finally get table 3, it appears AFER=0%,MSE=0, thus we stop to calculate. And get $D_q(0.000003,0.000003)$ is the best time series forecasting model in SD which simulate and forecast the enrollment data of the University of Alabama.

TABLE I. APPLYING $D_q(0.0003,0.0003)$ TO FORECAST THE ENROLLMENT DATA OF THE UNIVERSITY OF ALABAMA

Year	Enrollments V_q	Difference W_q	Forecast D_q	V_q-D_q	$(V_q-D_q)^2$	$\lvert V_q-D_q\rvert/V_q$
1971	13055	-	-	-	-	-
1972	13563	508	13563	0	0	0.000000
1973	13867	304	13867	0	0	0.000000
1974	14696	829	14696	0	0	0.000000
1975	15460	764	15461	-1	1	0.000065
1976	15311	-149	15311	0	0	0.000000
1977	15603	292	15603	0	0	0.000000
1978	15861	258	15861	0	0	0.000000
1979	16807	946	16804	3	9	0.000178
1980	16919	112	16919	0	0	0.000000
1981	16388	-531	16387	1	1	0.000061
1982	15433	-955	15429	4	16	0.000259
1983	15497	64	15497	0	0	0.000000
1984	15145	-352	15142	3	9	0.000198
1985	15163	18	15163	0	0	0.000000
1986	15984	821	15973	11	121	0.000688
1987	16859	875	16859	0	0	0.000000
1988	18150	1291	18150	0	0	0.000000
1989	18970	820	18970	0	0	0.000000
1990	19328	358	19324	4	16	0.000207
1991	19337	9	19337	0	0	0.000000
1992	18876	-461	18869	7	49	0.000371
AFER						0.0097%
MSE					10.5714	

TABLE II. APPLYING $D_q(0.00003, 0.00003)$ TO FORECAST THE ENROLLMENT DATA OF THE UNIVERSITY OF ALABAMA

Year	Enrollments V_q	Difference W_q	Forecast D_q	$V_q - D_q$	$(V_q - D_q)^2$	$\lvert V_q - D_q \rvert / V_q$
1971	13055	-	-	-	-	-
1972	13563	508	13563	0	0	0.000000
1973	13867	304	13867	0	0	0.000000
1974	14696	829	14696	0	0	0.000000
1975	15460	764	15460	0	0	0.000000
1976	15311	-149	15311	0	0	0.000000
1977	15603	292	15603	0	0	0.000000
1978	15861	258	15861	0	0	0.000000
1979	16807	946	16807	0	0	0.000000
1980	16919	112	16919	0	0	0.000000
1981	16388	-531	16388	0	0	0.000000
1982	15433	-955	15433	0	0	0.000000
1983	15497	64	15497	0	0	0.000000
1984	15145	-352	15145	0	0	0.000000
1985	15163	18	15163	0	0	0.000000
1986	15984	821	15983	1	1	0.000063
1987	16859	875	16859	0	0	0.000000
1988	18150	1291	18150	0	0	0.000000
1989	18970	820	18970	0	0	0.000000
1990	19328	358	19328	0	0	0.000000
1991	19337	9	19337	0	0	0.000000
1992	18876	-461	18875	1	1	0.000053
AFER						0.0006%
MSE					0.0952	

TABLE III. APPLYING $D_q(0.000003, 0.000003)$ TO FORECAST THE ENROLLMENT DATA OF THE UNIVERSITY OF ALABAMA

Year	Enrollments V_q	Difference W_q	Forecast D_q	$V_q - D_q$	$(V_q - D_q)^2$	$\lvert V_q - D_q \rvert / V_q$
1971	13055	-	-	-	-	-
1972	13563	508	13563	0	0	0.000000
1973	13867	304	13867	0	0	0.000000
1974	14696	829	14696	0	0	0.000000
1975	15460	764	15460	0	0	0.000000
1976	15311	-149	15311	0	0	0.000000
1977	15603	292	15603	0	0	0.000000
1978	15861	258	15861	0	0	0.000000
1979	16807	946	16807	0	0	0.000000
1980	16919	112	16919	0	0	0.000000
1981	16388	-531	16388	0	0	0.000000
1982	15433	-955	15433	0	0	0.000000
1983	15497	64	15497	0	0	0.000000
1984	15145	-352	15145	0	0	0.000000
1985	15163	18	15163	0	0	0.000000
1986	15984	821	15984	0	0	0.000000
1987	16859	875	16859	0	0	0.000000
1988	18150	1291	18150	0	0	0.000000
1989	18970	820	18970	0	0	0.000000
1990	19328	358	19328	0	0	0.000000
1991	19337	9	19337	0	0	0.000000
1992	18876	-461	18876	0	0	0.000000
AFER						0.0000%
MSE					0.0000	

III. CONCLUSION

For a time series, selecting the best time series model in SD by using automatic optimization search method and let the MSE=0 meanwhile AFER=0%. The $D_q(0.000003,0.000003)$ is the best time series forecasting model of the enrollment data of University of Alabama 1971~1992. The fact that the prediction accuracy of the existing fuzzy time series prediction model is not very high has been fundamentally improved.

ACKNOWLEDGEMENTS

This research was financially supported by Sanya Academy of science and technology cooperation project 2016YD04, 2015YD33, Hainan Provincial Natural Science Foundation 714283.

REFERENCES

[1] Qiu Wan-ren, Liu Xiao-dong. “Surver on theory and application of fuzzy time series forecasting models”. Fuzzy System and Mathematics, 28(3), pp. 173-181, 2014

[2] G. Udny Yule. "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers" [J].Philosophical Transactions of the Royal Society. A: Mathematical, Physical and Engineering Sciences, 1927, Vol.226, No.636-646, pp267-298

[3] Q Song, B S Chissom. Fuzzy series and its models. Fuzzy Sets and Systems, Vol. 54, pp.269-277, 1993

[4] Q Song, B S Chissom. Forecasting enrollments with fuzzy time series—Part I.Fuzzy Sets and Systems, Vol.54, pp. 1-9, 1993

[5] L A Zadeh. Fuzzy set .Fuzzy Sets and Systems, Vol. 8, pp. 338-353, 1965

[6] M Stevenson, J Porter. Fuzzy time series forecasting using percentage change as the universe of discourse. Proceedings of World Academy of Science, Engineering and Technology, Vol. 55, pp. 154-157, 2009

[7] V R Uslu, E Bas, U Yolcu, E Egrioglu. “A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations”, Swarm and Evolutionary Computation, vol.15, pp. 19-26, 2014.

[8] S.R. Singh, “A simple method of forecasting based on fuzzy time series”, Appl. Math. Comput, vol.186, pp. 330–339, 2007.

[9] C.H. Cheng, T.L. Chen, H.J. Teoh, C.H. Chiang, “Fuzzy time series based on adaptive expectation model for TAIEX forecasting (MEPA)”, Expert Syst. Appl. vol. 34, pp.1126–1132, 2008.

[10] T A Jilani, S M A Burney. “M-factor high order fuzzy time series forecasting for road accident data”, IEEE-IFSA 2007, World Congress, Cancun, Mexico, June 18-21, Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007.

[11] C.H. Cheng, T.L. Chen, H.J. Teoh, C.H. Chiang, “Fuzzy time series based on adaptive expectation model for TAIEX forecasting (TFA)”, Expert Syst. Appl. Vol.34, pp.1126–1132, 2008.

[12] R C Tsaur, J C O Yang, H F Wang. “Fuzzy relation analysis in fuzzy time series model”, Comput. Math. Appl. vol. 49, pp. 539–548, 2005.

[13] Bhagawati P. Joshi and Sanjay Kumar., A computationalmethod for fuzzy time series forecasting based on difference parameters. International Journal of Modeling, Simulation, and Scientific Computing, vol.4, no.1, pp. 1250023-1-1250023-12, 2013.

[14] T A Jilani, S M A Burney, C Ardil.“ Multivariate high order fuzzy time series forecasting for car road accidents”, International Journal of Computational Intelligence, vol. 4, no. 1, pp. 15-20, 2007.

[15] C. Kai, F Fang-Ping, C Wen-Gang. “A novel forecasting model of fuzzy time series based on K-means clustering’’, IWETCS, IEEE, pp. 223-225, 2010.