Interval Prediction and Stability Analysis of Time Series (Part Ⅰ: Theory)

Xintao Xia[1, a], Zhen Chang[1, b*], Yunfei Li[1, c], Bin Liu[1, d] and Liang Ye[1, e]

[1]Mechatronical Engineering College, Henan University of Science and Technology, Luoyang 471003, China

[a]xiaxt1957@163.com, [b]zhenc1992@163.com, [c]631617926@qq.com, [d]lbin1992@foxmail.com, [e]172682823@qq.com

*The corresponding author

Keywords: Time series; Grey bootstrap method; Interval prediction; Fuzzy-set theory; Stability

Abstract. Time series of different performance attributes are produced in the process of runtime of products, and any time series contain a large amount of information about the system evolution, so the information of the future evolution can be extracted from the time series to make a forecast or stability analysis. In this paper, based on the grey bootstrap method to establish a grey bootstrap distribution of time data sequences, the interval prediction of performance signal can be obtained by the given confidence level; Then according to the fuzzy-set theory, the fuzzy similar relation of engineering practice is changed into the fuzzy equivalence relation of space vector, and the stability analysis of time series is acquired by the given λ threshold. Two sets of models can effectively assess the change trend and performance evolution signs of time series, helping us to timely grasp the work performance situation.

Introduction

Time series for a certain performance attribute of product are formed by the order of time measurement value. According to the analogy or extension of the development process, the direction, trend and dynamic running level of time series can be achieved by predicting or analyzing next period time. If we can make full use of the temporal information, the fault diagnosis of product performance can be effectively completed [1-2]. The information mining of time series is attached highly attention by academia and engineering field, especially the aerospace, finance, economy, astronomy and geology [3-6]. While the early analysis models of time series are all almost linear, at present, more and more found that the nonlinear models can reasonably explain the practical engineering problem with the increasing requirement of the product quality indicators [7].

The range of time series is directly related weather performance characteristics can full play to its during working periods. Time series are usually accompanied by inherent evolution rule of product, so we can extract useful information to analyze the stability variation or to predict product performance weather is in good standing in the future. Based on nonlinear time series of performance data, the interval forecasting model is established by the grey bootstrap method [8], to analyze the range situation of interval evolution and promptly identify wave information of product performance signal; Then based on the fuzzy-set theory to analyze the stability of time series [9], evolution signs of time data sequences are evaluated by segment handling the raw data, and the stability of the product during operation is comprehensively discussed.

The Interval Prediction of Time Series

The Grey Bootstrap Prediction Model of Time Series. Suppose the vector of time series is expressed as

$$X = (x(1), x(2), ..., x(n), ..., x(N)) \tag{1}$$

where $X(n)$ is the nth data of the raw data; N is the number of the data in X. The first bootstrap samples ψ_1 is obtained by an equiprobable sampling N times with replacement from Eq. (1). And B simulation samples can be acquired by repeating B times in a row as follows:

$$\Psi = (\Psi_1, \Psi_2, ..., \Psi_b, ..., \Psi_B) \tag{2}$$

where $\boldsymbol{\psi}_b$ is the bth sample of bootstrap samples and can be expressed as

$$\Psi_b = (\psi_b(1), \psi_b(2), ..., \psi_b(n), ..., \psi_b(N)) \tag{3}$$

According to the grey system theory, suppose the first-order accumulated generating operator (1-AGO) of $\boldsymbol{\psi}_b$ is given by

$$\Gamma_b = (\gamma_b(1), \gamma_b(2), ..., \gamma_b(n), ..., \gamma_b(N)) \tag{4}$$

$$\gamma_b(n) = \sum_{k=1}^{n} \psi_b(k) \tag{5}$$

The grey prediction model can be given by

$$\frac{\mathrm{d}\gamma_b(n)}{\mathrm{d}n} + c_1\gamma_b(n) = c_2 \tag{6}$$

where c_1 and c_2 are the coefficients to be estimated.

Suppose the vector of mean series

$$Z_b = (z_b(2), z_b(3), ..., z_b(n), ..., z_b(N)) \tag{7}$$

$$z_b(n) = (0.5\gamma_b(n) + 0.5\gamma_b(n-1)) \tag{8}$$

where n=2,3,…, N.

The least-squares solution to Eq. (6) with the initial condition $\gamma_b(1)= \psi_b(1)$ is

$$\hat{\gamma}_b(n+1) = \left(\psi_b(1) - \frac{c_2}{c_1}\right)\exp(-c_1 n) + \frac{c_2}{c_1} \tag{9}$$

where the coefficients c_1 and c_2 are given by

$$(c_1, c_2)^{\mathrm{T}} = (D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}\Psi_b^{\mathrm{T}} \tag{10}$$

$$D = (-Z_b, I)^{\mathrm{T}} \tag{11}$$

$$I = (1,1,...,1) \tag{12}$$

According to the inverse AGO, the prediction value of n+1th for time series is given by

$$\hat{\psi}_b(n+1) = \hat{\gamma}_b(n+1) - \hat{\gamma}_b(n) \tag{13}$$

There will be B data in the n+1 moment, as follows:

$$Y = \hat{\psi}_b(n+1) \qquad b=1,2,\ldots,B \tag{14}$$

Due to B is very big, a probability density function of time series X can be obtained as follows:

$$F_w = f(x) \tag{15}$$

where F_w is the grey bootstrap frequency function or grey bootstrap distribution.

Interval Estimation. When the significance level $\alpha \in [0, 1]$, the confidence level can be given by

$$P = (1-\alpha) \times 100\% \tag{16}$$

The interval estimation is obtained by the probability density function in Eq. (15) under the condition of the confidence level is P, as follows:

$$[X_L, X_U] = [X_{\alpha/2}, X_{1-\alpha/2}] \tag{17}$$

where X_L is the lower limit of the estimated interval; X_U is the upper limit of the estimated interval.

The Stability Analysis of Time Series

The Fuzzy Equivalence Relation and Its Significance. The fuzzy equivalent relation should be used when analyzing system stability, but in the engineering practice, the fuzzy similar relation is often obtained, therefore it is necessary to transform the fuzzy similar relation into the fuzzy equivalence relation. Fortunately, the fuzzy equivalence relation of time series is rightly obtained by using the transitive closure method in the fuzzy-set theory, and the solving method as follows.

The time series data can be divided into m groups, viz. the original data X can be divided into m samples:

$$Z_i = \left(Z_{i1}, Z_{i2}, ..., Z_{ik}, ..., Z_{in}\right) \qquad k=1,2,\ldots,n, \qquad i=1,2,\ldots,m \tag{18}$$

A set is given by

$$\hat{Z} = \left(Z_1, Z_2, ..., Z_t, ..., Z_m\right) \tag{19}$$

where m is the number of samples; n is the sample size of each sample; Z_{ik} stands for the kth data of the ith sample.

For any fuzzy relation R, if present

$$T(R) = R^{h-1} = R^h = ... \quad h=1,2,3,\ldots \tag{20}$$

where $T(R)$ is the fuzzy equivalence relation and can be calculated according to Eq. (20), the steps as follows:

The first step is to calculate $R^2=R \circ R$;

The second step is to calculate $R^4=R^2 \circ R^2$;

…

Until $R^{2q}=R^q$ is obtained at the qth step, the process will be stopped.

where "$\circ$" is the fuzzy operation $M(\wedge,\vee)$ of matrixes; "$\vee$" stands for "or" operation to maximize; "$\wedge$" stands for "and" operation to minimize. Such as A=(0.7,0.4,0.2), B=(0.3,0.6,0.4), then $A \circ B^T$=(0.7$\wedge$0.3)$\vee$(0.4$\wedge$06)$\vee$(0.2$\wedge$0.4)= 0.3$\vee$0.4$\vee$0.2=0.4.

$\boldsymbol{R}^q$ is the fuzzy equivalent relation of $T(R)$, as follows:

$$T(R) = R^q \tag{21}$$

and

$$T(R) = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1l} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2l} & \cdots & \alpha_{2m} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \cdots & \alpha_{3l} & \cdots & \alpha_{3m} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{i1} & \alpha_{i2} & \alpha_{i3} & \cdots & \alpha_{il} & \cdots & \alpha_{im} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \alpha_{m2} & \alpha_{m3} & \cdots & \alpha_{ml} & \cdots & \alpha_{mm} \end{bmatrix} = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1l} & \cdots & \alpha_{1m} \\ & 1 & \alpha_{23} & \cdots & \alpha_{2l} & \cdots & \alpha_{2m} \\ & & 1 & \cdots & \alpha_{3l} & \cdots & \alpha_{3m} \\ & & & & \vdots & & \vdots \\ & & & & 1 & \cdots & \alpha_{im} \\ & & & & & & \vdots \\ & & & & & & 1 \end{bmatrix} \tag{22}$$

with $0 \le \alpha_{il} \le 1$.

$$\alpha_{il} = \begin{cases} 1, & i = l \\ \alpha_{li} & i \ne l \end{cases} \tag{23}$$

where α_{il} is the fuzzy equivalent relation of time series between the ith sample and the lth sample; viz. it is the conform degree of characteristics between Z_i and Z_l, or the coefficient of fuzzy equivalence relation having following significances:

(1)The closer to 1 α_{il} is, the better conform degree of characteristics between Z_i and Z_l is, viz. the smaller variation characteristics between the two samples is.

(2)The closer to 0 α_{il} is, the worse conform degree of characteristics between Z_i and Z_l is, viz. the more obvious variation characteristics between the two samples is.

(3) In particular, when α_{il}=1, Z_i and Z_l are exactly the same with no system error; When α_{il}=0, Z_i

and Z_l is irrelevant with significant variations.

So the diagnosis and analysis of the stability for time series can be realized by the above theory.

In practical engineering, α_{il} =1 and α_{il} = 0 are almost inexistent. Based on the concept of fuzzy mathematics, the evolution situation of research objects can be diagnosed by the optimal level λ and the level cut sets A_λ.

The Foundation of Stability Evolution. In fuzzy-set theory, the 0 and 1 indicate two extreme states of things between true and false, 0 indicating two research objects unrelated, and 1 indicating two entities closely related or relationship absolutely clear, so λ level and λ level cut sets A_λ can be used to diagnose the significant stability variation.

if

$$\alpha_{il} > \lambda \tag{24}$$

Then there is no stability variation between Z_i and Z_l under the λ level;

if

$$\alpha_{il} \le \lambda \tag{25}$$

Then there will be stability variation between Z_i and Z_l under the λ level;

In fuzzy-set theory, λ determines an entity's border from one extreme to another, viz. the threshold. When λ = 0.5, its fuzzy character reaches the peak, both true and false. $\lambda \ge 0.5$ means the relationship is gradually clear between Z_i and Z_l with a high similarity and no variation [10]. So the λ can be determined to be 0.5 in practical judgment.

The Fuzzy Characteristics of Stability Variation. The set of stability coefficient is defined as

$$U = \left(u_1, u_2, ..., u_j, ..., u_{m-1}\right) \tag{26}$$

where

$$u_j = \frac{\sum_{i=1}^{m-j} \alpha_{i,i+j}}{m-j} \qquad u_j \in [0,1],\ j=1,2,...,m\text{-}1 \tag{27}$$

where u_j stands for the stability coefficient, viz. the subsection mean value of fuzzy equivalence relation α_{il}; m is the sample size; j is the order of sampling time for each sample, viz. the time parameter. According to the stability coefficient u_j to judge the performance evolution signs of time series, $u_j > \lambda = 0.5$ means the product keep a good running; $u_j < \lambda = 0.5$ indicates that the stability is down during operation with obvious change trend, and it is likely to cause the performance damage of products.

In conclusion, the interval prediction of time series can better reflect its signal fluctuations during the work; In addition, the stability coefficient of time series can timely identify the evolution laws of product performance. Therefore the proposed grey bootstrap method can effectively predict the range state of product performance, and the stability coefficient of fuzzy relation can accurately monitor the performance evolution situation of time series.

Summary

(1) Based on grey bootstrap and fuzzy stability coefficient methods, the interval prediction and stability analysis model of time data series are proposed to predict variation range and analyze stability of product performance under the nonlinear conditions that the performance parameters, priori information, probability density and trend change of research objects are all unknown, then the state estimation of product performance components is realized during their operation.

(2) Combining the change trend of prediction intervals and stability coefficient, the evolution process of performance can be real-time evaluated. The hidden danger of failure can be timely found and serious accidents can be avoided.

(3) There is no the best model for the actual experimental data, only the most suitable model, so structuring a new analysis model of time series is the key issue of scholars and experts.

Acknowledgements

This project is supported by National Natural Science Foundation of China (Grant Nos. 51075123, 50375011 and 50675011).

References

[1] I. Antoniano-Villalobos and S.G Walker: Journal of Time Series Analysis, Vol. 63 (2016), p.3.

[2] Z.C. Zhang, G. Li and G.F. Chu: Vehicle System Dynamics, Vol. 53 (2015) No.6, p.833.

[3] H.J. Sadaei, R.Enayatifar and F.G. Guimaraes: Neurocomputing, Vol. 175 (2016), p.782.

[4] T. Guyet and H. Nicolas: Pattern Recognition Letters, Vol. 70 (2015), p.17.

[5] M. Matsumoto and A. Ikeda: Journal of Remanufacturing, Vol. 5 (2015) No.1, p.1.

[6] A.S. Lysyak and B.Y. Ryabko: Problems of Information Transmission, Vol. 52 (2016) No.1, p.92.

[7] J.F.L.D. Oliveira and T.B. Ludermir: Neurocomputing, Vol. 180 (2016), p.27.

[8] X.T. Xia, X.Y. Chen and Y.Z. Zhang: Measurement, Vol. 41 (2008) No.6, p.687.

[9] X.T. Xia and L. Chen: Measurement, Vol. 46 (2013) No.3, p.1349.

[10] X.T. Xia, Z.Y. Wang and Y.S. Gao: Measurement Science and Technology, Vol. 11 (2000) No.4, p.430.