ISSN: 2664-9640

A Feature Extraction technique Based on Static Divided Symbol Sequence Entropy

Zhenyi Chen, Xiong Hu

School of Shanghai Maritime University, Shanghai 201306, China.

Abstract

Gearbox is an important transmission equipment of quay crane hoisting mechanism. In order to accurately extract the degradation features from the vibration monitoring signal, a degradation feature extraction technique based on the static divided symbol sequence entropy is proposed. Considering the uniformity of the symbolization standard, the technique takes the root mean square of the signal in health condition as the basis, and combines the scale coefficient to establish a uniform basic scale. At the same time, the symbol set is expanded to enhance the information content and the ability of characterizing complexity of signal in large-value region. The Logistic chaotic sequence and the lifetime signal of hoisting mechanism gearbox are used for analysis respectively. The results show that the proposed technique is able to characterize the complexity of the nonlinear time series, and sensitively describe the performance degradation of the hoisting mechanism gearbox. The calculation speed is fast, which will lay a method foundation for further evaluating the health condition of large-scale quay crane in the port.

Keywords

Quay crane, Degradation feature extraction, Entropy, Symbol sequence entropy, Gearbox.

1. Introduction

The quay crane is a widely used large-scale port hoisting machinery, which mainly completes loading and unloading functions for port containers. Hoisting reduction box is a key transmission component of the hoisting mechanism and its running condition will directly affects the safety and reliability performance. The quay crane usually works in harsh natural environment and special working condition, therefore, the vibration monitoring signal of hoisting reduction box presents nonlinear, non-stationary and non-periodic characteristic. At the same time, the vibration signal is also mixed with large number of noise and shock components under working conditions, which increases the difficulty of vibration signal analysis[1]. In addition, the quay crane has a long liftime and harsh working environment, so it is difficult to monitor online and verify the technique's effectiveness. Therefore, how to extract the degradation features that characterize the operating condition from the complex and special monitoring signals is of great research significance for accurately assessing the health condition of quay crane and implementing Condition-Based Maintenance (CBM) [2].

As the basis of health condition evaluation, the purpose of degradation feature extraction is to mine performance degradation principle contained in the signal quantitatively, In view of the nonlinear, non-stationary and non-periodic characteristics of vibration monitoring signals, in recent years, signal analysis methods based on complexity theories including information entropy, fractal dimensions and chaos had been widely used in degradation features extraction for rotating machinery such as rolling bearings and gears. Some techniques had been proposed such as fuzzy entropy [3], permutation entropy [4], dispersion entropy [5], amplitude spectrum entropy [6-7], fractal dimension [8-9] and so on. On account of the difficulty of collecting

lifetime vibration monitoring signals, accelerated test data from IMS [10] and IEEE PHM 2012 [11] are usually introduced for verification. Compared with the accelerated lifetime signals above, vibration signals of the quay crane is more complex, and some effective methods verified in accelerated test data are proved out of work when processing the vibration signals from the quay crane.

Symbol dynamics is a signal symbolization theoretical method which has been successfully applied in fields of network security [12], biomedicine [13], and mechanical engineering [14]. This method is used by means of digitizing the original time series and extracting the main trend. At the same time, the noise components can be reduced and thereby greatly improving the effect and calculation efficiency. The combination of symbol dynamics and complexity theory is able to effectively extract the complexity quantitatively and characterize complex nonlinear systems.

Based on the above analysis, in order to effectively extract the degradation features of the quay crane hoisting gearbox, combining the symbol dynamics theory and information entropy, on the basis of the existing Basic Scale Entropy, a degradation feature extraction based on static divided symbol sequence entropy is proposed. The analysis and verification are carried out with the Logistic sequence and the lifetime dataset from the gearbox respectively. The paper is organized as follows: Section 2 briefly introduces Basic Scale Entropy and Section 3 proposed the new technique named static divided symbol sequence entropy. The simulated analysis of Logistic sequence is proposed in Section 4. In Section 5, the proposed method is verified and the results are discussed. Finally, the conclusion of this paper is given in Section 6.

2. Theory of Basic Scale Entropy

Basic Scale Entropy (BSE) is a typical feature analysis method based on static symbol division. It is mainly used in ECG signal processing in medical field [15]. In recent years, it has also some preliminary applications in field of mechanical equipment fault diagnosis [16]. The principle of the basic scale entropy method is as follows:

Supposing u is a one-dimensional time series with length N. Transforming the sequence in phase space:

$$X(i)=[u(i),u(i+L),\ldots,u(i+(m-1)L)] \tag{1}$$

Among them, m is the transformation dimension, L is the delay factor which demands $i+(m-1)L\leq N$. Generally, $L=1$, and the structure of matrix X is: $(N-m+1)\times m$. Later, symbolizing each m-dimensional vector and converting it to a sequence S with m-dimensional vector symbols.

$$S_i(X_i)=\{s(i),s(i+L),\ldots,s(i+(m-1)L)\} \tag{2}$$

In the formula above, $s\in A$:A={0,1,2,3}, and the principle of symbol symbolization are as follows.

$$S_i(X_i)=\begin{cases} 0: \bar{u}<u_{i+k}\leq \bar{u}+a\times BS \\ 1: u_{i+k}>\bar{u}+a\times BS \\ 2: \bar{u}-a\times BS<u_{i+k}\leq \bar{u} \\ 3: u_{i+k}\leq \bar{u}-a\times BS \end{cases} \tag{3}$$

Among them, $\bar{u}$ and BS represent the mean and basic scale of the ith m-dimensional vector, respectively, and BS is defined as follows:

$$BS(i)=\sqrt{\frac{\sum_{j=1}^{m-1}[u(i+j)-u(i+j-1)]^2}{m-1}} \tag{4}$$

In the formula above, the arguments a is used as basic scale parameter, which needs to be properly selected in practical applications. This method uses $a\times BS$ as the symbolization standard.

Counting and calculating the symbol pattern distribution probability P in sequence S. The total pattern of the four symbols is $\pi=4^m$, and the probability of the symbol pattern in S is calculated as follows:

$$p(\pi)=\frac{Num\{t\,|\,(u_t,u_{t+1},...,u_{t+m-1})\ hastype\ \pi\}}{N-m+1} \tag{5}$$

Among them, $1\le t\le N-m-1$, and Num represents the number.

Finally, the basic scale entropy BSE is calculated as follows:

$$H(m)=\frac{-\sum p(\pi)\log_2 p(\pi)}{\log_2(4^m)} \tag{6}$$

According to the principle of basic scale entropy, this method is able to describe the fluctuation pattern of the sequence. The larger the value, the more complex the fluctuation pattern, and vice versa.

3. Proposed of Static Divided Symbol Sequence Entropy

The basic scale entropy method takes $a\times BS$ as the symbolization standard to quantitatively measure the fluctuation pattern, and each series needs to calculate the basic scale BS according to formula (4). This will bring about two shortcomings, firstly, the inconsistency of the basic scale means the difference of the symbolization standard, and it is difficult to uniformly measure the changing of sequence complexity. Secondly, the length of time sequence will affect calculation speed, and the value of the parameter a will affect the symbolization standard.

Based on the above considerations, a feature extraction method of Static Divided Symbol Sequence Entropy (DSSE) is proposed. The principle is as follows:

Step1: Assuming that u is a one-dimensional time series with length N, converting u into an m-dimensional vector X.

$$X(i)=[u(i),u(i+L),...,u(i+(m-1)L)] \tag{7}$$

Step2: Symbolizing each m-dimensional vector in turn and converting it into an m-dimensional vector symbol sequence S.

$$S_i(X_i)=\{s(i),s(i+L),...,s(i+(m-1)L)\} \tag{8}$$

where $s\in A$:A={0,1,...,K}, the conversion principle is as follows:

$$S_i(X_i)=\begin{cases}1,|u(i)|\le BS_0\\ 2,BS_0<|u(i)|\le 2\times BS_0\\ \cdots\\ j,(j-1)\times BS_0<|u(i)|\le j\times BS_0\\ \cdots\\ K,|u(i)|>K\times BS_0\end{cases} \tag{9}$$

Among them, K means the number of symbols, the symbols set is A={1,2,3,4,5} when K=5. BS_0 is the basic scale of symbolization, and this parameter adopts a fixed value. In this paper, this parameter is set as the product of the root mean square value of the first group of health condition signal and the coefficient a as follows.

$$BS_0=a*\mathrm{rms}(u_1),\quad 0<a<2 \tag{10}$$

In this method, K represents the size of the symbols set that is the number of symbolized regions. The base scale representation represents the size of the symbolized area.

Step3: Counting the symbol pattern distribution probability P in S. The arrangement pattern of K symbols is $\pi=K^m$, and the calculation method is the same as formula (5).

ISSN: 2664-9640

Step4: Calculating the entropy *DSSE* of the static divided symbol sequence entropy as follows:

$$H_0(m) = -\sum p(\pi)\log_2 p(\pi) \tag{11}$$

In order to compare and analyze the influence of parameters selection on the results, the calculation of sequence entropy is not normalized in this paper.

4. Simulated analysis of Logistic sequence

Logistic model, also known as 'Pest model' [17], is a typical second-order recursive polynomial, $x_t+1=\lambda.x_t(1-x_t)$, which is mainly used in nonlinear dynamics and chaos analysis. The bifurcation diagram of the Logistic model (the initial value is 0.4, the sequence length is 2000) is shown in Figure 1. When λ is greater than 3.446, the Logistic model begins to oscillate, and when λ is approximately equal to 3.567, it begins to enter a chaotic state. As λ increases, the complexity of the sequence also increases but it also includes non-chaotic sequences, such as the islands of stability phase between λ=3.829 and 3.86 when the sequence returns from the chaotic state to the initial oscillating state.

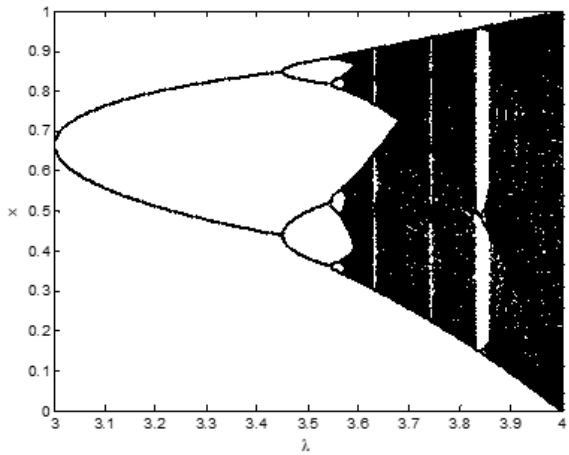


Fig.1 Logistic model bifurcation diagram

The Logistic model is processed by the proposed static divided symbol sequence entropy, where a=0.2, the effective value of the first group of sequences is 0.666, so the basic scale BS_0= a*0.666. The number of symbols is set as K=6, and the DSSE degradation feature curve is shown in Figure 2. It is obvious that the DSSE curve effectively characterize the process of model complexity from low to high, and at the same time, it is also sensitive to the initial oscillation at λ=3.446 and the islands of stability phase, indicating that this method is able to effectively characterize the complexity of the nonlinear time series.

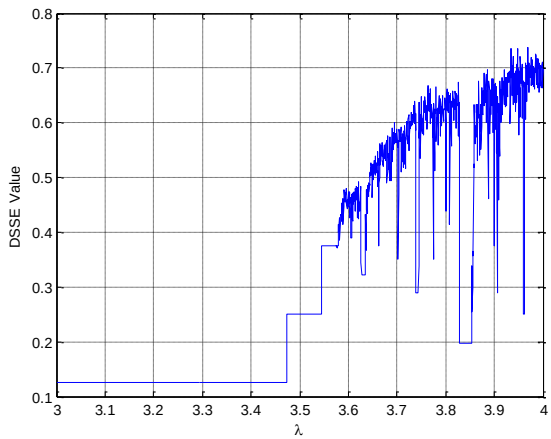
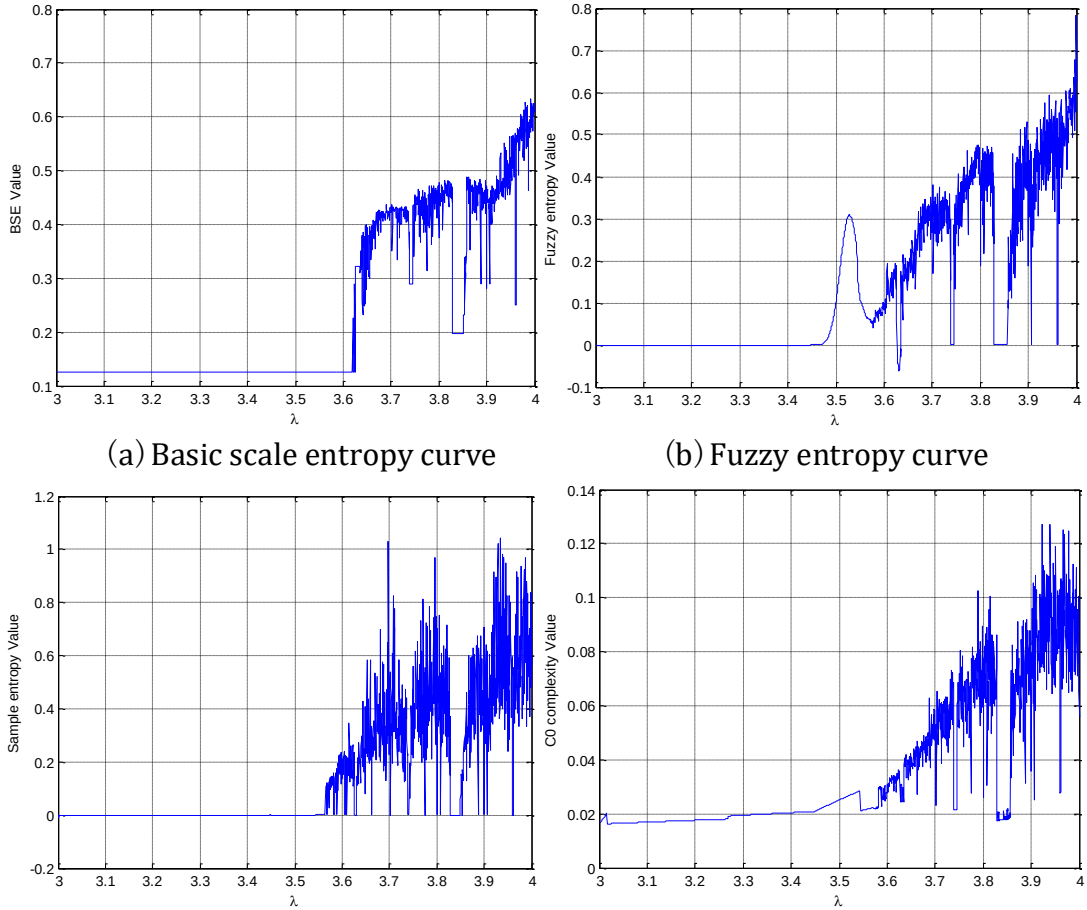


Fig. 2 DSSE degradation feature sequence of Logistic model

A horizontal comparison is made using typical complexity calculation methods, including Basic Scale Entropy [16], Fuzzy Entropy [18], Sample Entropy [19], and C0 complexity [20], the

ISSN: 2664-9640

degradation curves of the four methods are shown in Figure 3. As to the four methods, the overall upward trend is well reflected during the evolution process, but the Basic Scale Entropy and Sample Entropy curves fail to reflect the initial oscillation of the curve, and the complexity of Fuzzy Entropy in the initial oscillation process appears 'distorted'.



(a) Basic scale entropy curve (b) Fuzzy entropy curve

(c) Sample entropy curve (d) C0 complexity curve

Fig.3 Complexity based features degradation curves of Logistic model

The parameters and computational performance of the above methods are shown in Table 1. The operating environment is Intel Xeon® CPU E5-2450L, Memory 16.0GB, Matlab 2014a. In terms of operation time, C0 complexity is the fastest, while both Fuzzy Entropy and Sample Entropy are slower. The main reason lies that the C0 complexity is a typical structural complexity and only involves Fourier transform and statistical analysis, so it is the fastest in calculation. Fuzzy Entropy and Approximate Entropy are two typical behavioral complexities, due to the phase space transformation involved in the calculation, the calculation amount is large and the calculation time is long. The phase space transformation is also included in calculation process of DSSE technique, the amount of calculation is greatly reduced however due to the symbolization process. Therefore, the operation speed and structural complexity are almost the same order of magnitude. Therefore, the proposed static divided symbol sequence entropy is able to reflect the changing trend of signal complexity, and the operation is accurate and the speed and time are fast. It is suitable for degeneration feature extraction of mechanical equipment.

Table 1 Typical complexity degradation features setting and calculation time

Methods	Parameters setting	Time (second)
Basic Scale Entropy	m=4,a=0.5	0.8931

Static Divided Symbol Sequence Entropy	BS0 =0.2*rms(u1)，u1 is the first group signal	0.8433
Fuzzy Entropy	m=4, r= 0.2*std(s), n=10	32.5798
Sample Entropy	m=4, r= 0.2*std(s)	10.9290
C0 Complexity	a=0.8	0.3937

5. Summary

In order to extract the degradation feature of quay crane hoisting gearbox, a degradation feature extraction method based on static divided symbol sequence entropy is proposed and the technique is validated with simulation and lifetime vibration signals. The following conclusions are obtained.

(1) The proposed static divided symbol sequence entropy technique improves the basic scale entropy method, the basic scale of symbolization is unified, the symbols number variable and basic scale coefficient are added meanwhile to control the symbolic regions more flexibly. The verification of Logistic model and lifetime loading spectrum shows that this technique is able to sensitively describe the complexity of nonlinear time series. The higher the complexity, the larger the value, thus it is convenient to track the progress of equipment performance degradation.

(2) The symbols number variable and basic scale coefficient are added in this technique, the two parameters are able to change symbol pattern of the symbol sequence, thus affecting the value of DSSE. Among them, the increase of symbols number will gradually increase the information expression ability in high-amplitude region, improving the complexity resolution of the high-amplitude signal region. In addition, the basic scale coefficient is proposed to control the accuracy of symbol sequence's ability in signal information expressing.

References

[1] Hou Meihui, Hu Xiong, Wang Bing,etc. Degradation feature extraction for the ship-to-shore crane turning point based on Weibull distribution, JOURNAL OF VIBRATION AND SHOCK, 2019,38(22):198-203+243.

[2] Jagtap H P , Bewoor A K , Kumar R . Failure analysis of induced draft fan used in a thermal power plant using coordinated condition monitoring approach: A case study, Engineering Failure Analysis, 2020, 111:104442.

[3] Tang Zhanjun, Sun Runfa. Fan Bearing Fault Diagnosis Based on Multi-scale Fuzzy Entropy and STOA-SVM,Motor and Control Application,2021,48(12):66-70.

[4] Wang Gongxian, Zhang Miao, Hu Zhihui, etc.Bearing fault diagnosis based on multi-scale mean permutation entropy and parametric optimization SVM, JOURNAL OF VIBRATION AND SHOCK, 2022,41(01):221-228.

[5] Wang Mian, Liu Yong. Fault Diagnosis Method of Rolling Bearing Based on Time-shifted Multi-scale Dispersion Entropy and SVM, Machine Design and Research,2021,37(05):83-87.

[6] [6]Peng Ruitao, Luo Xiuqiang, Luo Yue, etc.Rolling Bearing Fault Diagnosis Based on Information Entropy and Alpha Stable Distribution, Machine Design and Research,2021,37(04):92-98.

[7] Chen Mang, Yu Dejie, Gao Yiyuan. Fault diagnosis of rolling bearings based on graph spectrum amplitude entropy of visibility graph, JOURNAL OF VIBRATION AND SHOCK, 2021,40(04):23-29.

[8] Du Xiaolei, Chen Zhigang, Wang Yanxue, etc.Application of morphological empirical wavelet transform and IFractalNet in bearing fault identification. Journal of Vibration Engineering, 2021,34(03):654-662.

[9] Xi Maosong, Xu Xin, Pan Hongxia. Early Fault Diagnosis of Supply and Delivery System based on k Value Optimization for VMD and Matrix Fractal, Machine Design and Research,2020,36(02):208-211+215.

[10] Lee Jay, Qiu Hai, Yu Gang,et al. Bearing dataset from IMS of university of Cincinnati and NASA ames prognostics data repository,http://ti.arc.nasa.gov/tech/dashpcoe/prognostic-data-repository.

[11] NECTOUX P,GOURIVEAUR,MEDJAHER K, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests, IEEE International Conference on Prognostics and Health Management. Denver,CO,2012.

[12] Yang Baowang. Low-rate-denial-of-service attack detection by symbolic dynamics method,Jounal of XIDIAN University,2018,45(01):140-144.

[13] Xin Yi, Zhao Yizhang, Mu Yuanhui. ECG Feature Analysis Based on Poincare Plot and Symbolic Dynamic, Transactions of Beijing Institute of Technology, 2017,37(10):1084-1089.

[14] Ding Chuang, Feng Fuzhou, Zhang Bingzhi, etc. MMSDE and its application in feature extraction of a planetary gearbox, JOURNAL OF VIBRATION AND SHOCK,2020,39(13):97-102+147.

[15] Zhao Tingting, Duan Xiaoli, Yan Bige. Multiscale base-scale entropy analysis of heart rate variability signals in the young and the elderly,Chinese Journal of Medical Physics,2019,36(01):88-91.

[16] Wang Bing, Hu Xiong, Li Hongru, etc. Rolling bearing performance degradation state recognition based on basic scale entropy and GG fuzzy clustering, JOURNAL OF VIBRATION AND SHOCK, 2019,38(05):190-197+221.

[17] Bryson B, Kunze H , Torre D L , et al. A Generalized Multiple Criteria Data-Fitting Model With Sparsity and Entropy With Application to Growth Forecasting. IEEE Transactions on Engineering Management, 2021, (99):1-12.

[18] Li Y , Wang S , Yang Y , et al. Multiscale symbolic fuzzy entropy: An entropy denoising method for weak feature extraction of rotating machinery. Mechanical Systems and Signal Processing, 2022, 162(7):108052.

[19] Sidorenko A V, Saladukha M A. Evaluation of operator's state under the influence of electromagnetic noise generator. System analysis and applied information science, 2021(4):45-53.

[20] Wang Wei, Hu Xiong,Wang Bing,etc.Identification of Degradation State for Bearing Performances Based on C0 Complexity and GG Fuzzy Clustering, Bearing, 2019(12):51-57.