Sensors & Transducers, Vol. 172, Issue 6, June 2014, pp. 124-129

Sensors & Transducers

Research on Health State Perception Algorithm of Mining Equipment Based on Frequency Closeness

[1] Gang Wang, [1] Yanjun Hu, [2] Guoqing Ji

[1] The Engineering Research Centre of Perception Mine, China university of Mine & Technology, Xuzhou, 221008, China

[2] Department of Aviation Ammunition, Air Force Logistics College, Xuzhou, China, 221006

[1] Tel.: 051683899723, fax: 051683899597

E-mail: wgcumt@yeah.net

Received: 16 April 2014 /Accepted: 30 May 2014 /Published: 30 June 2014

Abstract: The health state perception of mining equipment is intended to have an online real-time knowledge and analysis of the running conditions of large mining equipments. Due to its unknown failure mode, a challenge was raised to the traditional fault diagnosis of mining equipments. A health state perception algorithm of mining equipment was introduced in this paper, and through continuous sampling of the machine vibration data, the time-series data set was set up; subsequently, the mode set based on the frequency closeness was constructed by the δ neighborhood method combined with the TSDM algorithm, thus the forecast method on the basis of the dual mode set was eventually formed. In the calculation of the frequency closeness, the Goertzel algorithm was introduced to effectively decrease the computation amount. It was indicated through the simulation test on the vibration data of the drum shaft base that the health state of the device could be effectively distinguished. The algorithm has been successfully applied to equipment monitoring in the Huoer Xinhe Coal Mine of Shanxi Coal Imp&Exp. Group Co., Ltd. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Mining equipment, The dual mode set, Health state perception, Time-series data mining, Frequency closeness.

1. Introduction

The health state perception of mining equipment is intended to have an online real-time knowledge and analysis of the running conditions of large mining equipments, and to determine whether they are in their normal conditions. What's different from the fault diagnosis lies in that the failure modes of the equipment could be eliminated from our consideration and by selecting the appropriate mode base, the device modes can be classified into two types, i.e. the normal one and the fault one, which could cleverly avoid the adverse condition of the unclear equipment failure modes.

The vibration data of the machine is essentially time-series data, concretely, in Reference [1, 2], a time-series data mining (TSDM) method was presented, and for the transient set it proposed, the genetic algorithm was introduced in Reference [3] to search for the optimal transient set; in Reference [4], the frequency fuzzy closeness was introduced to measure the distance between the time series, based on which the level of similarity of the time-series data was further analyzed; and in Reference [5] and [6], the ant colony algorithm and Kurtogram algorithm were utilized for fault classification, respectively. The algorithms above are suitable for traditional fault diagnosis. However, due to the unknown prior failure modes of mining equipments, and combined with the health state perception idea, the concept of the dual mode set was introduced in this paper, and through continuous sampling of the machine vibration data, the time-series data set was set up; subsequently, the mode set based on the frequency closeness was constructed by the δ neighborhood method combined with the TSDM algorithm, thus the forecast method on the basis of the dual mode set was eventually formed. In the calculation of the frequency closeness, the Goertzel algorithm was introduced to effectively decrease the computation amount.

The first part of this paper analyzes the similarities and differences between the health state perception of mining equipment and traditional fault diagnosis. The second part gives two kinds of guidelines for determining similarity between two time series. The third part studies the forecast method on the basis of the dual mode set. The fourth part provides the TSDM algorithm. The fifth part gives the simulation and the sixth part makes the conclusion. It was indicated through the simulation test on the vibration data of the drum shaft base that the health state of the device could be effectively distinguished.

2. Similarity Criteria of Time series

Let $X = \{x_1, x_2, ..., x_n\}$ be a time series, wherein $x_1, x_2, ..., x_n$ are the measured values at the time of $t_1, t_2, ..., t_n$, respectively; and $|X|$ be the length of the measurement vector $X = \{x_1, x_2, ..., x_n\}$; $A = \{X_1, X_2, ..., X_n, ...\}$ be the measured time-series set to be judged, and then the similarity classification of the time series could be presented by the following formalized definition:

Give the time series *X*, the time-series set *A* to be judged, the similarity measurement function $sim()$ and the similarity judgment algorithm $alg()$, find the series set *B* similar to *X* from the time-series set *A*, i.e.

$$B = \{X_i \in A \mid sim(alg(X, X_i))\} \quad (1)$$

2.1. Euclidean Distance Similarity Criteria

The Euclidean distance is a most common discrimination method of the time-series distance. Give two time series of $X = \{x_1, x_2, \cdots x_n\}$ and $Y = \{y_1, y_2, \cdots y_n\}$, and their Euclidean distance is defined as:

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

There exist the following disadvantages when Euclidean distance is employed for the cluster analysis of the time series:

1) The lengths of the time series to be judged must be equal [4];

2) It is easy to obtain a wrong clustering result.

That is because the amplitude of the vibration signal is characterized by instantaneous changes, and two waveforms without similar amplitudes usually have similar vibration characteristics. For time-domain waveforms, the determinations of the average value, the maximum peak, the kurtosis, the intensity and other parameters are generally to be conducted.

For this, the similarity discrimination method of Euclidean distance does not apply to the time-domain similarity of the machine vibration signals, thus the transform-domain method shall be employed.

2.2. Frequency Closeness Similarity Criteria

The frequency characteristic changes are often resulted from the operating condition variation and the failure occurrence of the device, consequently, so as to better observe the dynamic behavior of the machine running status, frequency conversion could be conducted on the signals to obtain the frequency-domain information. Frequency-domain analysis is one of the most widely applied signal processing methods in the condition diagnosis of rotating machines. The frequency spectrum is a generic term, which is generally composed of the amplitude spectrum and the power spectrum, etc., and the analysis effects of both are the same when the phase information is ignored [7]. In the upper computer, the frequency spectrum analysis is achieved by fast Fourier transformation.

Give a time series of $X = \{x_1, x_2, ..., x_n\}$, with the length measured as |*X*| = *N*, whose Fourier transform pairs are

$$X_k = \sum_{i=0}^{N-1} x_i W_N^{ik} \quad 0 \le k \le N-1, \quad (3)$$

$$x_i = \frac{1}{N} \sum_{k=0}^{N-1} X_k W_N^{-ik} \quad 0 \le i \le N-1, \quad (4)$$

where $W_N = e^{-j\frac{2\pi}{N}k}$ is the twiddle factor.

To use base-2 FFT algorithm, it requires the FFT signal sampling points is general an integer power of 2 [8].

The similarity judgment of two time series can be realized by comparing the approximation degree of the frequency domains. To this two sequences Euclid frequency Closeness is defined as:

$$N(X,Y)=\frac{1}{n}\sqrt{\sum_{k=1}^{n}(X_k-Y_k)^2}\ , \qquad (5)$$

where X_k and Y_k are the Fourier transform coefficients of and, respectively, i.e. the frequency peaks included.

It could be seen from the definition of closeness that the closer X and Y are, the closer is $N(X,Y)$ to 0.

3. Mathematical Analysis Method of Time-series Similarity

3.1. Mode Cluster and Mode Set

A mode is a manifestation pattern of the characteristic quantity in the time series, and any mode hasn't been determined to be 'health' or 'failure' is called uncertain mode, which include the past and the present un-determined conditions.

For the actual un-determined time series, their uncertain modes are not completely the same; in order to describe the attributions of each uncertain mode, the concepts of mode cluster and mode set were both introduced. The mode cluster P is a compact set composed of all the modal space points whose distances to the known mode p are less than δ, i.e. $P=\{a\in R^Q \mid d(p,a)\le\delta\}$, where d is the spatial distance, and Q is the dimension of the modal space.

The set composed of all the mode clusters is called the mode set, denoted by C. The series belonging to different mode clusters but the same mode set is considered to be in the same state. As a consequence, the mode set is the smallest division for the final state discrimination. A two dimensions mode cluster and mode set are shown in Fig. 1, where N is the spatial distance.

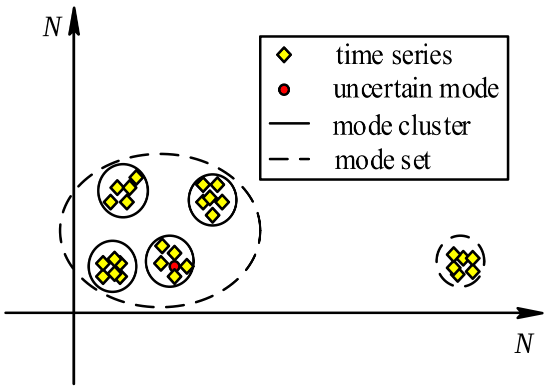

Fig. 1. Mode cluster and mode set.

3.2. Discrimination Steps of Time-series Similarity

1) Embed the time series X into a Q-dimensional real space, with FFT transformation adopted, a reconstructed state space was established.

2) Define an event description function $g(X)$, which represents the future event value of the state space X. To be simplified, it is practical to take the value of 1 to indicate the event occurrence, while the value of 0 to indicate no event occurrence.

3) Define the objective function f, which reflects the event description ability of the mode cluster P. The optimal mode clusters could be obtained through the optimal solutions of the objective function f, all of which could then constitute the mode set.

4. Algorithm Achievement based on TSDM

4.1. Algorithm Flowchart

Algorithm flowchart of TSDM [2] is shown as Fig. 2.

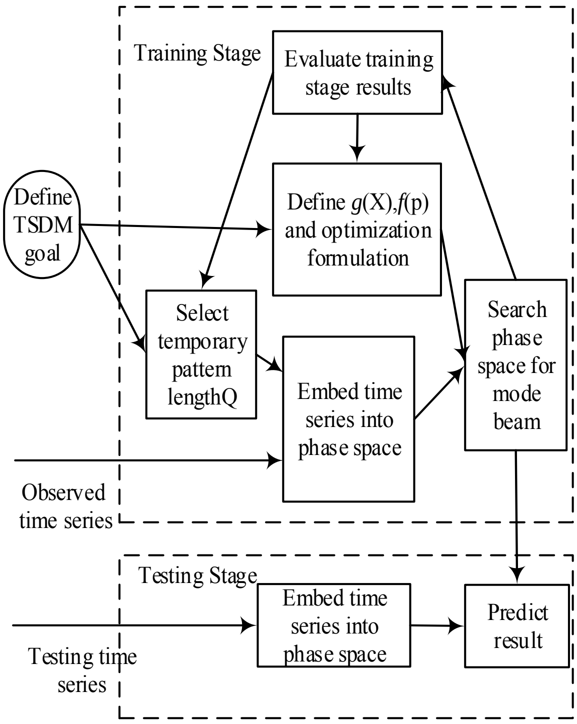

Fig. 2. Algorithm flowchart of TSDM.

4.2. Determinations of the Event Description Function and the Objective Function

For the time series $X=\{x_1,x_2,...,x_n\}$, the event description function g(X) is to describe the fault event occurring condition of the time series X, because of the unknown prior failure modes of the mining equipment, the device modes can be classified into the normal and the failure modes, i.e. two kinds of mode sets, for this, define the event description function as

$$g(X)=\begin{cases}1, & X \text{ is the fault state}\\ 0, & X \text{ is the normal state}\end{cases} \tag{6}$$

Let the mode cluster $P=\{a \mid d(p,a)\le \delta_1\}$, and define the objective function

$$f(p)=\min_{a\in P} d(p,a) \tag{7}$$

And $d(p,a)=N(P_k,A_k)$, that is the frequency closeness of the two series.

If $a\in C$, then $g(a)=0$, otherwise $a\in \overline{C}$, $g(a)=1$.

4.3. Simplified Formula and Schema Length Updating Algorithm

The frequency closeness is the Euclidean distance between the Fourier transformation coefficients of the time series $X=\{x_1,x_2,\cdots x_n\}$ and $Y=\{y_1,y_2,\cdots y_n\}$. Considering that the machine vibration frequency is associated with the rotating speed, which will only appear at some special frequency points, therefore, it's only necessary to know the DFT values of M special points, and $M << \log_2 N$. For this reason, with the Goertzel algorithm utilized, the DFT calculation could be expressed as a linear filtering operation with the periodicity of the twiddle factor.

Because $W_N^{-kN}=1$, $X_k=W_N^{-kN}X_k$ $=\sum_{m=0}^{N-1}x(m)W_N^{-k(N-m)}$.

Let $y_k(n)=\sum_{m=0}^{N-1}x(m)W_N^{-k(n-m)}$ and $h_k(n)=W_N^{-kn}u(n)$, and then $y_k(n)=x(n)*h_k(n)$. That is $y_k(n)$ could be treated as the convolution between the series $x(n)$ with finite length of N and the filter with the unit impulse response of $h_k(n)$, and the output of this filter at $n=N$ is just the DFT value at the frequency $\omega_k=\frac{2\pi}{N}k$, i.e.

$$X_k=y_k(n)|_{n=N} \tag{8}$$

The system function of the filter with the impulse response of $h_k(n)$ is

$$H_k(n)=\frac{1}{1-W_N^{-k}z^{-1}} \tag{9}$$

For this filter, there exists a pole at the frequency of $\omega_k=\frac{2\pi}{N}k$ on the unit circle. As a consequence, when a group of data to be transformed passes through *M* single-pole filter banks, the DFT values at the required points can be eventually calculated out.

The value of *M* can be determined by the quantity of the maximum values in the DFT spectrum, and the calculated value of M is to be the temporary schema length *Q*.

Assuming that the amplitudes corresponding to the selected *Q* poles is X'_k, calculate $N(X',Y')$, if

$$\frac{N(X',Y')}{N(X,Y)}>\delta_2 \tag{10}$$

Then Update *Q*, select the largest M-1 values among the maximums, and let *Q*=*M*-1, repeat the calculation above until $\frac{N(X',Y')}{N(X,Y)}<\delta_2$, and adopt the *Q* value of the last calculation as the final schema length. Reckon the characteristic frequency into the mode cluster to constitute the mode set.

4.4. Achievement of Early-warning

Judge the test data in accordance with the optimal mode set achieved as per the steps above, as long as the state-space point ranges within a certain optimal mode cluster of the mode set C, the device will be considered to be healthy, otherwise it may be faulty, when early-warning information will be sent out.

5. Simulation Test

5.1. Vibration Data Collection and Arrangement of the Drum Shaft Base

Actual vibration data collection and transmission of the drum shaft base from the Huoer Xinhe Coal Mine of Shanxi Coal Imp&Exp. Group Co., Ltd. were conducted, wherein 5 groups of time series were randomly selected from the running drum shaft base.

Since the effective values of each sample were different, the similarity level of the time series cannot be characterized through the Euclidean distance. Further, certain vibration failure features were exhibited in the frequency characteristics of the graphics, therefore, the similarity in the frequency domain can be determined through comparing the frequency-domain closeness.

The time domain and frequency domain graph of 5 samples are shown in Fig. 3. For ease of display and analysis, time domain waveform amplitude is normalized to maximum amplitude.

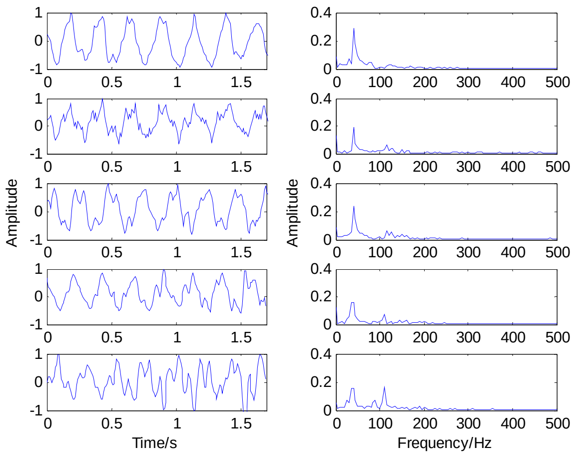

Fig. 3. Time domain samples and their Fourier transform.

5.2. Similarity Analysis Process of the Vibration-similar Series for the Drum Shaft Base

Select the measured data of Sample 1 as the time series that has been judged in healthy status, and calculate the Euclidean distances between Samples 2~5 and Sample 1, respectively, together with the frequency closeness as well, as shown in Table 1. It could be obtained from the calculation results that Sample 3 was close to Sample 1, which could also be considered as healthy, while Sample 4 and 5 were fault for their larger frequency closeness, and Sample 2 was in a pending status due to its unobvious data.

Table 1. Compare between Samples 2~5 and Sample 1.

Sample	Euclidean distances	Frequency closeness	Verdict
Sample 2	5.8302	0.1982	Pending
Sample 3	13.4506	0.1125	Similar
Sample 4	9.8406	0.2544	Dissimilar
Sample 5	10.7774	0.2934	Dissimilar

Adopt Sample 3 as the sample of the health database, and calculate the Euclidean distances and the frequency closeness between it and the rest in turns, as shown in Table 2. From the calculation results, Sample 2 was close to Sample 3, which could be considered as healthy, while Sample 4 and 5 were surely fault for their frequency closeness were still larger.

Table 2. Compare between Samples 2,4,5 and Sample 3.

Sample	Euclidean distances	Frequency closeness	Verdict
Sample 2	10.0001	0.1318	Similar
Sample 4	9.0756	0.2086	Dissimilar
Sample 5	8.8915	0.2544	Dissimilar

The formation process of the mode set was shown in Fig. 4, wherein Sample 2 was initially in a pending status, but finally belonged to the healthy mode set.

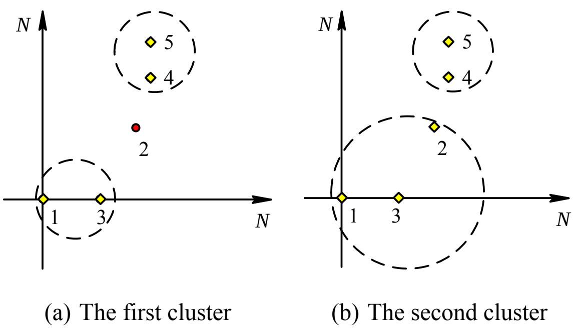

Fig. 4. The formation process of the mode set.

In the calculation above, the frequency closeness was still obtained after the DFT calculations of all the points, with a larger amount of computation, while in the actual implementation, the optimization formula and the schema length upgrading algorithm could be employed, wherein at the initial stage, the fundamental frequency, 2~5 times the frequency, the power frequency 50 Hz and its multiplications can be adopted to conduct the frequency closeness calculations. In this experiment, according to the test, two frequencies, i.e. 40 Hz and 110 Hz, can be selected to achieve discrimination. One time calculation result is shown in Table 3.

Table 3. Compare between Samples 2~5 and Sample 1 (two frequency points are selected).

Sample	Euclidean distances	Verdict
Sample 2	0.1008	Pending
Sample 3	0.0512	Similar
Sample 4	0.1470	Dissimilar
Sample 5	0.2058	Dissimilar

The schema lengths Q after optimization shall be selected in accordance with different measuring objects. The vibration data of the belt in the coal preparation plant and the sieve bearing were shown in Fig. 5 and Fig. 6, in general, Q=5-10 could be chosen for accurate calculation.

It could be seen from the analysis and calculation results above that this algorithm can effectively distinguish the two states of 'health' and 'failure', if any index is further defined, the possibility scale of failure can also be determined. After the schema length *Q* was optimized, the computation got faster with less calculation amount, consequently, it is favorably suitable for the online health monitoring and discrimination for the operating status of mining equipments.

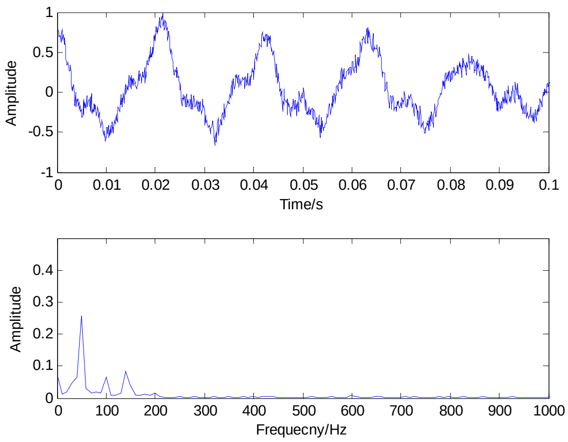

Fig. 5. Vibration data of the belt.

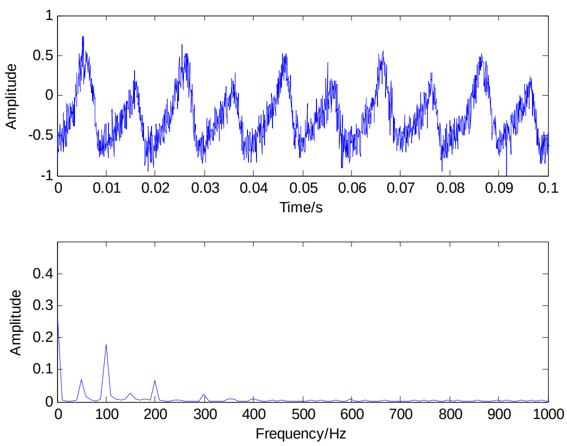

Fig. 6. Vibration data of the sieve bearing.

6. Conclusions

A health state perception algorithm of mining equipment based on TSDM was introduced in this paper. At first, the time series similarity theory is introduced and time series analysis methods are described in detail. A TSDM-based algorithm is given in the following. In the calculation of the frequency closeness, the Goertzel algorithm was introduced to effectively decrease the computation amount. Simulation result shows the efficient of the method. The algorithm has been successfully applied to equipment monitoring in the Huoer Xinhe Coal Mine of Shanxi Coal Imp&Exp. Group Co., Ltd.

Acknowledgements

This paper is supported by "The Fundamental Research Funds for the Central Universities (2010QNB29)" and "National Science and Technology Support sub-topics (2013BAk06B05, 2012BAH12B02)".

References

[1]. Povinelli Richard J., Feng Xin, Temporal pattern identification of time series data using pattern wavelets and genetic algorithms, *Artificial Neural Networks in Engineering*, 1998, pp. 691-696.
[2]. Povinelli Richard J., Using genetic algorithms to find temporal patterns indicative of time series events, in *Proceedings of the GECCO Workshop: Data Mining with Evolutionary Algorithms*, 2000, pp. 80-84.
[3]. Su Shengchao, Zhang Zhengdao, Zhu Daqi, Fault prediction for rotating machinery based on time series data mining, *Journal of Nanjing University of Aeronautic & Astronautics*, 38, 7, 2006, pp. 120-123.
[4]. Bai Liang, Wang Han, Li Hui, Luo Xingqi, Vibration fault diagnosis based on time-series similarity mining for hydropower units, *Journal of Hydroelectric Engineering*, 29, 6, 2010, pp. 229-236.
[5]. Wu Dongmin, Shao Jianping, Rui Yannian, Research on CNC machine fault diagnosis based on ant colony algorithm and neural network, *Machinery Design & Manufacture,* 1, 2013, pp. 165-167.
[6]. Wang Hongchao, Chen Jin, Dong Guangming, Cong Feiyun, Application of resonance demodulation in rolling bearing fault feature extraction based on fast computation of kurtogram. *Journal of Vibration and Shock*, 32, 1, 2013, pp. 35-37.
[7]. Wang Jinfu, Li Fucai, Signal processing methods in fault diagnosis of machinery — analyses in frequency domain, *Noise and Vibration Control*, 1, 2013, pp. 173-180.
[8]. Wang Yanfen, Digital signal processing: principles and practice, *Tsinghua University Press*, Beijing, 2013.