Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):2335-2340

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Fault prediction of fan bearing using time series data mining

Xingjie Chen and Wenfa Zhu

School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai, China

---

ABSTRACT

The fault symptoms are regarded as a sort of temporal patterns hidden in a time series. A novel method based on time series data mining is proposed for the prediction of fan bearing fault. The time series, which is formed by large numbers of fan bearing vibration data, is embedded into a reconstructed phase space with time-delay. In this phase space, Genetic Algorithms are used to search for the optimal temporal pattern clusters which are the criteria to identify temporal patterns. The optimal collection of temporal pattern clusters is then used to test the other bearing vibration data of fan. Once the symptoms are detected, the fault is forecasted. The simulation results show the method is efficient.

Key words: time series, data mining, fan, bearing, fault prediction

---

INTRODUCTION

Fan is a sort of rotating machinery which takes rotors and other rotative parts as its main body for work. In many large-scale industries, it is the core of equipments for production. If the fault of fan happens, it will not only hamper the normal work but also bring on needless loss. So fault prediction of fan is beneficial for the production and safety of enterprise.

At present, there are few researches on fault prediction for rotating machinery, and the primary method for fault prediction is the method based on time series prediction. Typical time series prediction methods generally use linear model to approximate data series, and they are only effective for fault prediction in linear system. However, actual rotating machinery system is a nonlinear system, the typical methods are unfeasible for it. With the nonlinear mapping capability of neural networks, the time series prediction method based on neural networks has been used to forecast fault of nonlinear system in recent years. Tse and Atherton used recurrent neural networks to forecast fan fault of a chemical plant in Hongkong [1]; Zhang and his fellow used a combined neural networks model to predict the fault of steam turbine [2]; Other people used BP neural networks to prognose rotating machinery fault [3]. But these neural networks based prediction methods still exist some shortages: firstly, it is difficult yet to ascertain the framework of networks, sometimes it can but relies on experience; secondly, the error of extrapolating curve using neural networks is yet incapable of being analyzed.

Time series data mining (TSDM) is to extract some useful information or knowledge unknown previously but latent from large numbers of past and present time series data. Inspired by concepts in data mining and dynamical systems, in 1998, Richard J. Povinelli and Xin Feng presented the TSDM framework [4~5] which focuses on identifying temporal patterns for characterization and prediction of time series events. There are several significant features of the proposed method. First, the method focuses on the identification of the temporal patterns that are characteristic of the events. Second, with the temporal patterns identified, the new method focuses on event prediction rather than complete time series prediction. This allows the prediction of complicated time series events such as the fault events from a rotating machine. Third, the objective function in the optimization reflects the goal of the time series being examined, e.g., fault happens, and is problem specific.

This paper, which is divided into six sections, presents the fault prediction method based on TSDM. The second section introduces the fundamental TSDM method. The third section discusses the characteristic of fault events from a rotating machine and the formation of time series for fault prediction. The fourth section establishes the TSDM method for predicting fault events. The fifth section presents experimental results from predicting fan end bearing fault. The last section summarizes this paper.

1. Outline of the Time Series Data Mining Method[5]

Given a training time series $X = \{x_t, t = 1, \cdots, N\}$, the method is as follows:

Step 1. The time series X is unfolded into IR^Q—a reconstructed phase space, called simply phase space here—using time-delayed embedding. The unfolding mechanism maps X into IR^Q. Specifically, a set of Q time series observations $\{x_{t-(Q-1)\tau}, \cdots x_{t-2\tau}, x_{t-\tau}, x_t\}$ taken from X map to $\mathrm{x}_t=\{x_{t-(Q-1)\tau}, \cdots x_{t-2\tau}, x_{t-\tau}, x_t\}^T$, where x_t is a column vector or point in the phase space, τ is the time delay, and t is an integer in the interval $[(Q\text{-}1)\tau+1, N]$.

Step 2. In order to correlate a temporal pattern (past and present) with an event (future), a real valued function $g(\mathrm{x}_t)$, the so-called "event characterization function," is defined and associated with each phase space point x_t. The event characterization function represents the value of future "eventness" for the present phase space point x_t.

Step 3. Construct a heterogeneous (in the sense that Q may take multiple values) collection of temporal pattern clusters C, such that, C is the optimizer of the objective function f, where a temporal pattern cluster P is defined as a ball consisting of all points within a certain distance δ of a temporal pattern $\mathbf{p}$ in the aforementioned IR^Q phase space and the temporal pattern $\mathbf{p}$ is a $Q\times1$ vector in the same IR^Q phase space. The objective function f maps a collection of temporal pattern clusters C onto the real line, thereby providing an ordering to collections of temporal pattern clusters according to their ability to characterize events. The objective function f is constructed in such a manner that its optimizer C is predictive of the events of interest. An event is then predicted whenever a phase space point x_t formed from a set of Q time series observations $\{x_{t-(Q-1)\tau}, \cdots x_{t-2\tau}, x_{t-\tau}, x_t\}$ is within one of the temporal pattern clusters P that comprise C.

2. Establishment of the Time Series

Although there are many factors which result in faults of rotating machinery, the extrinsic representation of the factors is mostly from the vibration of mechanism whose performance parameters can be reflected by the vibration signals sensitively and directly. Besides, the evolution from original fault to the system fault is a gradual process that means the amplitude of the present vibration is related with the past vibration. The relationship between the past and present in the amplitude sequence is the basis of the forecast. So it is necessary to establish a time series to find the relationship. In our paper, the vibration data of mechanism components, which is generated before and after faults happened, is selected every time in order to establish the time series where the relationship aforementioned hides. Fig. 1 shows us a sample of the time series which is composed of the bearing vibration data of a certain fan, and the small amplitudes describe the normal state, and the larger amplitudes describe the fault state.

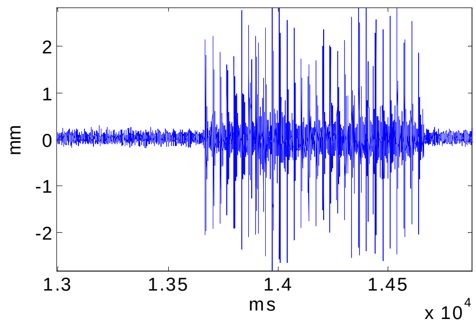

Fig.1: Time series sample of fan bearing vibration

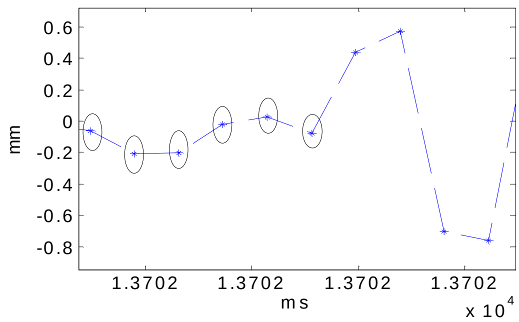

Fig.2: Fan bearing vibration time series before fault

4. Fault Prediction Based On TSDM

Given a training time series of bearing vibration $X = \{x_t, t = 1, \cdots, N\}$. The symptoms before the fault of machinery happens can be regarded as temporal patterns hidden in the time series, thus the TSDM method are used to identify them so as to achieve fault prediction. Fig. 2 shows us a fan bearing vibration time series before and when fault happens, the six ellipses in this picture represent a six dimension temporal pattern.

4.1 Selection of the Event Characterization Function g and Definition of the Objective Function f

Definition 1. The binary time series $Y = \{y_t, t = 1, \cdots, N\}$, when t is the first sampling time of fault happens y_t=1, or else y_t=0 (include *t* are the other fault time and the normal time), which indicates the events with a one indicating a fault event and a zero indicating a nonevent.

The event characterization function *g* represents the value of future "eventness" for the phase space point x_t, so for n-steps prediction, the fault event characterization function is

$$g(x_t) = y_{t+n\tau}, \qquad (1)$$

where τ is time delay . When $g(x_t)$ =1, x_t is a temporal pattern n-step before a fault event.

Definition 2. Let temporal pattern clusters $P=\{a\in \mathbf{IR}^Q : d(\mathbf{p}, a)\le \delta\}$ and *C* is the collection of temporal pattern clusters. True positive (t_p) is the number of fault events within *P* or *C*; false positive (f_p) is the number of nonevents within *P* or *C*; true negative (t_n) is the number of nonevents outside *P* or *C*; false negative (f_n) is the number of fault events outside *P* or *C*.

There are two primary objects for fault prediction, one is the minimal probability of failing to predict fault and the other object is the maximal prediction accuracy. So two objective functions are defined separately as follows:

$$f_1(C) = \frac{f_n}{t_p + f_n} \qquad (2)$$

$$f_2(P_i) = \frac{t_p}{t_p + f_p} \qquad (3)$$

According to definition 2, the first objective function in formulation (2) is used to determine the efficacy of a collection *C* of temporal pattern clusters in total probability of failing to character or predict fault; the second objective function in formulation (3) can be used to represent the characterization or prediction accuracy of a temporal pattern *P*. Obviously, the optimal formulation min$f_1(C)$ and max$f_2(P_i)$ ($\forall P_i \in C$) are used to achieve the two objects of fault prediction. In order to find a minimal set of temporal pattern clusters that is a optimizer of the first objective function, the optimal formulation min$f_1(C)$ is subject to min$c(P_i)$, where $c(P_i)$ is the number of P_i which comprises the set of temporal pattern clusters *C* .

4.2 Selection of the Phase Space Dimension Q

The value of *Q*, i.e., the length of the temporal pattern **p** and the dimension of the reconstructed phase space, is selected based on Takens' Theorem [8], which states that if $Q\ge 2m+1$, where *m* is the original state space dimension, the reconstructed phase space is guaranteed to be topologically equivalent to the original state space. However, there are some difficulty in estimating *m* for the time-delay embedding process. Estimating *m* is more difficult when the original time series contains both stochastic and deterministic signals since the stochastic component may require that *m* be infinite. Fortunately, as shown in [4], [9], [10], [11], useful information can be extracted from the reconstructed state space even if its dimension is less than 2*m*+1. So using the principle of parsimony, temporal patterns with small *Q* are examined first.

4.3 Search for a Single Optimal Temporal Pattern Cluster Using Genetic Algorithm

A variant of the well-known simple Genetic Algorithm is employed here to search for a single optimal temporal pattern cluster P_i^*. The objective function used by GA was presented in formulation (3) and a hash table [12] is used to store previously calculated fitness values, thereby achieving a computational speedup without sacrificing accuracy.

The phenotype for the GA, $P_i = [\mathbf{p} \ \ \delta]$, is encode as a binary string. The decoding of the genotype is defined as

$$p_i = \frac{p_{\max} - p_{\min}}{2^l - 1}\sum_{j=0}^{l-1} 2^j p_{i,j} + p_{\min} \qquad (4)$$

where l is the length of the gene used to encode p_i, $p_{\max} = \max X$, $p_{\min} = \min X$, and *X* is the training time series. The radius is defined as

$$\delta = \frac{\delta_{\max}}{2^l - 1} \sum_{j=0}^{l-1} 2^j \delta_j \qquad (5)$$

where $\delta_{\max} = Q(p_{\max} - p_{\min})$, Q is the dimension of **p** and the Manhattan distance is chosen as the metric for the reconstructed phase space.

A tournament of size two is used as the selection mechanism. Mutation in the range of 0-0.1% is used. The stopping criterion is convergence of the fitness values. Elitism of one is employed.

4.4 Search for an Optimal Collection of Temporal Pattern Clusters for Fault Prediction

Let $Q_{\min}$ and $Q_{\max}$ be the minimum and maximum time-delay embedding dimension, respectively. The function in formulation (2) is the objective function for the collection of temporal pattern clusters. We search for an Optimal Collection of Temporal Pattern Clusters using the following algorithm:

Set $Q = Q_{\min}$

a Unfold the training time series into the reconstructed IR^Q phase space and search for an optimal temporal pattern cluster P_i^* in this phase space using the GA described above. Define a threshold β to accept a temporal pattern cluster, if $f_2(P^*{}_i) > \beta$, then P_i^* is regarded as a temporal pattern cluster we need and repeat step a after removing the clustered phase space points from the phase space.

b Else if $f_2(P^*{}_i) \le \beta$, $Q = Q + 1$ and goto step a.

c While $Q > Q_{\max}$, the search stops. Evaluate the training results. If necessary, select the new range of Q and search again. An Optimal Collection of Temporal Pattern Clusters C^* is comprised of all the optimal temporal pattern clusters P_i^* we need.

4.5 Fault Prediction using the Optimal Collection of Temporal Pattern Clusters C^*

The optimal collection of temporal pattern clusters C^* for fault prediction is found by the search process in 4.4. An fault event is then predicted n-steps ahead whenever a phase space point x_t formed from a set of Q time series observations $\{x_{t-(Q-1)\tau}, \cdots x_{t-2\tau}, x_{t-\tau}, x_t\}$ is within one of the temporal pattern clusters P_i^* that comprise C^*.

5. Application—Fan Fault Prediction

In this section, our method is applied to the prediction of fan bearing fault. The fan bearing vibration data is from electronic engineering and compute science laboratory of Case Western Reserve the vibration data University and the data for experiment is sampled at 12KHZ. We mix the bearing vibration data when the rotating speed of motor is 1797rp/min, 1772 r/min, 1750 r/min, and 1730 r/min respectively, then use the mixed data to form a time series according to the third section in this paper. The time series is actually divided to two parts, the training part and the test part. The training time series comprised of 480,000 data contains 23 fault series and the other 480,000 data form the testing time series which contains 17 fault series. A sample of training time series can be seen in Fig.1.

5.1 Training Results

For one-step fault prediction, event characterization function $g(t) = y_{t+\tau}$; the range of phase space dimensions Q is [1,15] and τ is set to 1; a parameter set is used for the GA: the initial population size multiplier is 10, the population size is 50, the elite count is one, the gene length l is eight, the mutation rate is 0.05% and the convergence criterion is 0.6.

Table 1-The Parameter of the Optimal Collection of Temporal Pattern Clusters

Q	4	5	6	7	8	10
p	(−0.000822, −0.000719, −0.000411, −0.000206)	(0.0266, 0.0257, 0.0271, 0.0275, 0.0252)	(0.000514, 0.00298, 0.00154, 0.00412, 0.00206, 0.00452)	(0.0494, 0.0466, 0.0410, 0.0422, 0.0380, 0.0337, 0.0342)	(−0.0345, −0.0349, −0.0140, −0.0561, −0.0846, −0.0906, 0.0713, 0.0487)	(0.0176 ,0.0327, 0.0118, 0.0192, 0.0373, 0.0261, 0.0225, 0.0343, 0.0325, 0.0297)
δ	0.3338	0.6462	0.5147	0.2794	0.0111	0.1019

The results of the search in training stage are shown in Table 1 and Table 2. Six temporal pattern clusters whose dimension are 4, 5, 6, 7, 8, and 10 form the optimal collection employed to identify temporal patterns (fault symptoms). The probability of failing to character fault is 4.5%, and the prediction accuracy is 88%, which indicate the optimal collection has a good efficacy to character bearing fault.

Table 2-Characterization Results in training stage

Q	4	5	6	7	8	10	total
t_p	5	8	3	3	1	2	22
f_p	0	0	1	0	0	2	3
$f_n = 1$	Characterization accuracy	$f(P)=\frac{t_p}{t_p+f_p}=\frac{22}{25}$					88%
	Probability of failing to character fault	$f(c)=\frac{f_n}{t_p+f_n}=\frac{1}{23}$					4.5%

Table 3-Prediction Results In Testing Stage

Q	4	5	6	7	8	10	total
t_p	4	6	4	1	0	1	16
f_p	0	0	1	0	0	1	2
$f_n = 1$	prediction accuracy	$f(P)=\frac{t_p}{t_p+f_p}=\frac{16}{18}$					88.7%
	probability of failing to predict fault	$f(c)=\frac{f_n}{t_p+f_n}=\frac{1}{17}$					5.9%

5.2 Testing Results

The optimal collection discovered during the training process is applied to identify temporal patterns (fault symptoms) hidden in the testing time series. When a state value x_t formed from a set of Q testing data $\{x_{t-(Q-1)\tau}, \cdots x_{t-2\tau}, x_{t-\tau}, x_t\}$ is within one of the six temporal pattern clusters that comprise the optimal collection, x_t is regarded as a temporal pattern, i.e., a fault symptom value, and the bearing fault is then forecasted one-step ahead.

The results of test are seen in Table 3. The probability of failing to predict fault is 5.9%, and the prediction accuracy is 88.7%, which also indicate the feasibility of using the optimal collection to predict bearing fault.

5.3 Comparison of Results With TDNN

Here, we compare the above method with a time delay neural network (TDNN) [6]. The TDNN algorithm was provided with the same data set to train our prediction method, that is the previous 10 values of the vibration time series to predict the bearing fault event in next time step. Recall that the maximum dimension of any of the six temporal pattern clusters discovered in the training phase was 10. This indicates the number of previous values used in the prediction of a bearing fault.

Table 4- Prediction Results of the Two Methods

Stage	Training		Testing	
Method	Characterization Accuracy	Probability of failing to character fault	Prediction Accuracy	Probability of failing to predict fault
TDNN	86.2%	5.9%	78.7%	11.1%
Our Method	88%	4.5%	88.7%	5.9%

The TDNN has four layers with 10 neurons in the input layer, 20 neurons in the first hidden layer, 100 neurons in the second hidden layer, and one neuron in the output layer. Sigmoid style activations functions are used in the first three layers, and a threshold style is used in the output layer. The TDNN was trained for 1000 epochs.

Table 4 shows the results for our prediction method, and the TDNN. In comparing the results of the two method, it is clearly seen that our prediction method is superior to the TDNN.

CONCLUSION

The paper presented a fault prediction method based on TSDM, and this method is applied to fault prediction of fan bearing. Different from other fault prediction methods, the method predicts fault by mining latent temporal patterns in system, which can provide the current research of fault prediction in nonlinear system with a new approach.

REFERENCES

[1] Tse P W, Atherton D P. *Journal of Vibration and Acoustics*, 121(7): 355-362 , **1999**.
[2] ZHANG Zhiming, CHENG Huitao, XU Hong, HU Sangao. Neural Network Based Combining Prediction Model And Its Application in Condition Based Maintenance of Turbo-Generator Set. Proceedings of the Csee, 23(9), **2003**.
[3] WANG Fuliang. *Mining & Processing Equipment*, 33(11), **2005**.
[4] R.J. Povinelli and X. Feng. *Proc. Artificial Neural Networks in Eng. Conf.*, pp. 691-696,**1998**.
[5] Richard J. Povinelli. Using Genetic Algorithms to Find Temporal Patterns Indicative of Time Series Events. GECCO **2000** Workshop: Data Mining with Evolutionary Algorithms, 80-84.
[6] C.T. Lin and C.S.G. Lee, Neural Fuzzy Systems—A Neuro-Fuzzy Synergism to Intelligent Systems. Upper Saddle River, N.J.: Prentice-Hall, **2011**.
[7] R.J. Povinelli. Identifying Temporal Patterns for Characterization and Prediction of Financial Time Series Events. Proc. Temporal, Spatial and Spatio-Temporal Data Mining: First Int’l Workshop; revised papers, (TSDM ’00), pp. 46-61, **2000**.
[8] F. Takens. Detecting Strange Attractors in Turbulence. Proc. Dynamical Systems and Turbulence, pp. 366-381, **1980**.
[9] H.D.I. Abarbanel. Analysis of Observed Chaotic Data. New York: Springer, **1996**.
[10] E. Bradley. Analysis of Time Series. An Introduction to Intelligent Data Analysis, M. Berthold and D. Hand, eds., pp. 167-194, New York: Springer, **1999**.
[11] R.J. Povinelli and X. Feng. Data Mining of Multiple Nonstationary Time Series. Proc. Artificial Neural Networks in Eng., pp. 511-516, **1999**.
[12] R.J. Povinelli and X. Feng. Improving Genetic Algorithms Performance By Hashing Fitness Values. Proc. Artificial Neural Networks in Eng. Conf., pp. 399-404, **1999**.