

A method based on Dempster-Shafer theory and support vector regression-particle filter for remaining useful life prediction of crusher roller sleeve[★]

Haiping Liu^{1,*}, Jianjun Wu¹, Xiang Ye¹, Taijian Liao¹, and Minlin Chen²

¹ School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Gan Zhou 341000, PR China

² Faculty of Foreign Studies, Jiangxi University of Science and Technology, Gan Zhou 341000, PR China

Received: 13 April 2017 / Accepted: 30 September 2018

Abstract. In order to solve the problem of accurately predicting the remaining useful life (RUL) of crusher roller sleeve under the partially observable and nonlinear nonstationary running state, a new method of RUL prediction based on Dempster-Shafer (D-S) data fusion and support vector regression-particle filter (SVR-PF) is proposed. First, it adopts the correlation analysis to select the features of temperature and vibration signal, and subsequently utilize wavelet to denoising the features. Lastly, comparing the prediction performance of the proposed method integrates temperature and vibration signal sources to predict the RUL with the prediction performance of single source and other prediction methods. The experiment results indicate that the proposed prediction method is capable of fusing different data sources to predict the RUL and the prediction accuracy of RUL can be improved when data are less available.

Keywords: D-S theory / data fusion / RUL prediction / support vector regression / particle filter

1 Introduction

Roller sleeve as an important component is widely used in crusher, the status of its running performance directly affects the health of the whole equipment [1]. However, due to the complex working load, dust and other harsh working conditions, the service life of crusher roller sleeve is not long, and the accurate life of high speed spindle in all kinds of crushers is only thousands of hours. Once the working hours exceed the service life limit, the operation precision of roller will drop sharply and further cause that the machine cannot work properly. So it is very important to improve the reliability, safety and work efficiency of the crusher roller sleeve by means of prognostics and health management (PHM). It is an important part of PHM to predict the RUL of equipment and evaluate the performance of devices [2].

It is critical to establish an appropriate model in the life prediction process. In brief, condition-based monitoring is

becoming more and more significant, especially in the RUL prediction. Vibration signal on-line monitoring is one of the most effective methods to monitor the state of crusher roller sleeve health condition (SOH) [3]. The RUL prediction of crusher roller sleeve based on vibration monitoring data is divided into two steps: firstly, construct an indicator to accurately assess the performance degradation of crusher roll sleeve; subsequently, establish an effective model to predict the RUL of crusher roller sleeve.

How to establish an appropriate model under partially observable state is the key to predict the RUL accurately, and it is also the urgent demand for the industrial production. SVR-PF is such a machine learning algorithm to make classification and prediction under small samples [4]. This method based on the statistics theory has been successfully applied to the prediction in the financial, electric and other systems [5,6].

In this paper, a method using acceleration and temperature data is proposed firstly to solve the challenge of low RUL prediction precision based on single data source. However, the noise and vibration interferences caused by other mechanical and systems may severely obscure the roller sleeve signal collected from sensors and make it very challenging to reliably detect the effective components. For the above reasons, a variety of signal analysis methods have been proposed by researchers, such as time-domain, frequency-domain and time-frequency

[★] This work was supported by the National Natural Science Foundation of China under Grants 5136015 and 51665017 and the Science and Technology Research Programs of Science and Technology Commission Foundation of Jiangxi province under Grants 20142BBE50058 and 20161BBE80041.

* e-mail: 1834575793@qq.com

Table 1. BPA function combination based on D-S data fusion.

BPA function	$m_1(T)$	$m_1(a)$	$m_1(T \cup a)$
$m_2(T)$	$m_1(T)m_2(T)$	\emptyset	$m_1(T \cup a) m_2(T)$
$m_2(a)$	\emptyset	$m_1(a) m_2(a)$	$m_1(T \cup a) m_2(a)$
$m_2(T \cup a)$	$m_1(T) m_2(T \cup a)$	$m_1(a) m_2(T \cup a)$	$m_1(T \cup a)m_2(T \cup a)$

technique. Wavelet analysis is such a widely accepted approach. Then it selects the sensitive features of two signals as input and constructs the SVR-PF model to solve the problem that is difficult to predict with finite state data. The proposed method is evaluated using experimental data respectively. Finally, after assessing the prediction result errors, the conclusion is given in [Table 5](#).

2 Theory introduction of D-S data fusion and SVR-PF

2.1 Theory introduction of D-S data fusion

D-S theory fuses data from different sources through the basic probability assignment (BPA) function, and analyses the belief of all the possible propositions in the identification framework, so as to achieve the goal of data fusion [7–10].

To set the identification framework consists of evidence B and C , m_1 and m_2 are two BPA functions in the identification framework, $m_{1,2}$ is the fused BPA function, then the D-S data fusion can be expressed as:

$$m_{1,2}(\emptyset) = 0 \quad (1)$$

$$m_{1,2}(A) = \frac{1}{1-K} \sum_{B \cap C=A} m_1(B)m_2(C) \quad (2)$$

where K is the degree of conflict between the evidences B and C

$$K = \sum_{B \cap C=\emptyset} m_1(B)m_2(C). \quad (3)$$

D-S data fusion theory combines with the same view for different sources of the same problem and eliminates the all conflicting views at once, so that a more reliable fused posterior BPA function can be obtained.

The RUL prediction of crusher roll sleeve based on D-S data fusion has two data sources: (1) the RUL prediction based on temperature data; (2) the RUL prediction based on acceleration data. In this paper, the proposed prediction method based on D-S data fusion and SVR-PF fuses the results of two prediction methods to gain the fused RUL prediction result.

The whole identified framework is set as Ω

$$\Omega = \{T, a\}. \quad (4)$$

Because there's no intersection between T and a , prediction by acceleration data and prediction by temperature data are independent events, then the power set can be expressed as 2^Ω .

$$2^\Omega = \{\emptyset, \{T\}, \{a\}, \{T \cup a\}\}. \quad (5)$$

The meaning of all the propositions in the power set 2^Ω is explained as follows.

- $\{T\}$ represents the RUL prediction credibility obtained by temperature data;
- $\{a\}$ represents the RUL prediction credibility obtained by acceleration data;
- $\{T \cup a\}$ represents the RUL prediction credibility obtained by acceleration or temperature data.

Meanwhile, the BPA functions m_1 and m_2 defined in the power set 2^Ω mean that:

- m_1 represents the prediction credibility distribution obtained by temperature data in the power set 2^Ω ;
- m_2 represents the prediction credibility distribution obtained by acceleration data in the power set 2^Ω .

The combination of the BPA function based on data fusion is shown in [Table 1](#).

From [Table 1](#), the posterior BPA function based on data fusion can be expressed as:

$$m(T) = \frac{1}{1-K} \sum_{B \cap C=T} m_1(B)m_2(C) = \frac{b}{1-K} \quad (6)$$

$$m(a) = \frac{1}{1-K} \sum_{B \cap C=a} m_1(B)m_2(C) = \frac{c}{1-K} \quad (7)$$

where

$$b = m_1(T)m_2(T) + m_1(T \cup a)m_2(T) + m_1(T)m_2(T \cup a) \quad (8)$$

$$c = m_1(a)m_2(a) + m_1(T \cup a)m_2(a) + m_1(a)m_2(T \cup a) \quad (9)$$

$$K = \sum_{B \cap C=\emptyset} m_1(B)m_2(C) = m_1(a)m_2(T) + m_1(T)m_2(a). \quad (10)$$

That's the proposed RUL prediction method for crusher roller sleeve, it uses the posterior fusion BPA function and combines the prediction results of two prediction methods to gain a more accurate prediction result.

2.2 Basic theory of SVR-PF

2.2.1 Basic theory of particle filter

On the basis of the recursive Bayesian estimation [11–13], particle filter becomes a universal algorithm drawing samples from posterior distributions and assigns weights to all the particles by using the Monte Carlo method [13–16].

Particle filter has more excellent performance on nonlinear and non-Gaussian system than Kalman filter which only has good performance on linear and Gaussian system [17].

The particle filter system state space model can be described as:

$$\begin{cases} x_k = f(x_{k-1}, v_{k-1}) \\ z_k = h(x_k, n_k) \end{cases} \quad (11)$$

where x_k is the system state, z_k is either the system output or the measurement, v_{k-1} is the system noise, and n_k is the measurement noise.

We assume that the prior distribution $p(x_{0:k-1}^i | z_{1:k-1})$ of system is known and N samples from the posterior distribution of system (11) have drawn. The posterior distribution can be approximately described as:

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^N w_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (12)$$

where $\{x_k^i\}$ is the sample, $\{w_k^i\}$ is the sample weight, which have $\sum_i w_k^i = 1$. The higher is the weight, the higher is the sample probability. $\delta(\cdot)$ represents the Dirac-Delta function.

In order to solve the problem that is very difficult to sample directly from a posterior distribution, a good deal of the problem is the importance sampling technique. It can draw samples directly from the importance distribution. The importance distribution can be described as:

$$q(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^N \delta(x_{0:k} - x_{0:k}^i). \quad (13)$$

Plugging the importance distribution (13) into (12), then the weight can be updated:

$$\begin{aligned} w_k^i &= \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i) p(x_{0:k-1}^i | z_{1:k-1})}{q(x_k^i | x_{0:k-1}^i, z_{1:k}) q(x_{0:k-1}^i | z_{1:k-1})} \\ &= w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{0:k-1}^i, z_{1:k})} \end{aligned} \quad (14)$$

where $p(z_k | x_k^i)$ is the likelihood function, $p(x_k^i | x_{k-1}^i)$ is the state transfer distribution. If system (11) subjects to the Markov process, the weight update equation (14) can be reduced to:

$$w_k^i = w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)}. \quad (15)$$

We set state transfer distribution as the importance distribution:

$$q(x_k^i | x_{k-1}^i, z_k) = p(x_k^i | x_{k-1}^i). \quad (16)$$

If the likelihood function $p(z_k | x_k^i)$ and the prior weights are used to update the new weights [15], the weight renew equation can be reduced to equation (17):

$$w_k^i = w_{k-1}^i p(z_k | x_k^i). \quad (17)$$

There is a wider problem of PF, which is known as degeneracy phenomenon. In order to avoid the problem, resampling is a suitable method. If the system iterates without resampling, the weight of some particles will tend to zero, and all efforts for the weights calculation become meaningless.

The standard method to avoid the degeneracy phenomenon is to renormalize the distribution by removing the small weight particles and duplicating the large weight particles. The weights of all the particles are set to $1/N$ (N is the number of particles). The resampling algorithm of the standard PF is shown above.

$$N_{eff} = \frac{N}{1 + \text{var}(w_k^i)} \approx \frac{1}{\sum_{i=1}^N (w_k^i)^2} \quad (18)$$

where N_{eff} is the threshold of resampling.

2.2.2 Support vector regression-particle filter

The standard PF algorithm eliminates the small weight particles and duplicates the large weight particles to avoid the degeneracy phenomenon that would cause the loss of particle diversity. Which would make most particles aggregate around the larger weighted ones, so the degeneracy phenomenon still exists. In view of this problem, a new resampling algorithm known as SVR is introduced to rebuild a posterior distribution [18], which has an extremely fast learning speed and advantageous generalization capability. Moreover, SVR has a commendable performance in both classification and regression with a simple structure. Compared with other methods, SVR can avoid the degeneracy phenomenon and keep the diversity of particles in the case of limited samples. What's more, the training speed of SVR is much faster while obtaining better generalization. In view of these advantages, the SVR is selected in this paper to establish RUL prediction model. The application of the SVR is detailed in some studies [19,20].

The fundamental principle of SVR is known as an optimization problem expressed by a regularized functional with constraints [21], the form can be described as:

$$\begin{cases} \Omega = (f, f)_H \\ s.t. \sup_x |F(x) - F_l(x)| = \sup_x |F_l(x) - \int_{-\infty}^x f(t) dt| = \sigma_l < \varepsilon \end{cases} \quad (19)$$

where the regularized functional defined in Hilbert space and generated by σ_l is represented by $\Omega = (f, f)_H$. The error between the distribution functions $F(x)$ and their estimation $F_l(x)$ is represented by σ_l . The constraint is represented by ε .

The estimated probability density function (PDF) of distribution $F_l(x)$ is $F(x)$. Only the points $x_i (i=1, 2, \dots, m)$ in the particle set should be considered, so equation (19) can be reduced to:

$$\max_i \left| F_l(x) - \int_{-\infty}^x f(t) dt \right|_{x=x_i} = \sigma_l < \varepsilon. \quad (20)$$

If the PDF $f(x)$ is described by kernel functions:

$$f(x) = \sum_{i=1}^m \beta_i K(x_i, x) \quad (21)$$

Kernel function $K(x_i, x) = \varphi_r(x_i)\varphi(x)$ satisfies Mercer's condition. Then the regularized functional can be described as:

$$\Omega(f) = (f, f)_H = \sum_{i=1}^m \sum_{j=1}^m \beta_i \beta_j K(x_i, x_j). \quad (22)$$

The posterior distribution prediction can be described as an optimization problem with constraints:

$$\begin{cases} \min w_p(\beta) = \sum_{i=1}^m \sum_{j=1}^m \beta_i \beta_j K(x_i, x_j) \\ s.t. \max_i \left| F_i(x) - \sum_{j=1}^m \beta_j \int_{-\infty}^x K(x_j, t) dt \right|_{x=x_i} = \sigma_l. \end{cases} \quad (23)$$

Set $y_i = F_i(x_i)$, $w = [\beta_1, \beta_2, \dots, \beta_m]^T$, $z_j(x) = \int_{-\infty}^x K(x_j, t) dt$, $z_i = (z_i(x_1), z_i(x_2), \dots, z_i(x_m))$, ξ_i and ξ_i^* are non-negative slack variables, then equation (23) can be reduced to a quadratic programming problem:

$$\begin{cases} \min J(w, \xi_i, \xi_i^*) = \frac{1}{2} w^T w + C \left(\sum_{i=1}^m \xi_i + \sum_{i=1}^m \xi_i^* \right) \\ s.t. w^T z_i - y_i \leq \sigma_l + \xi_i \\ y_i - w^T z_i \leq \sigma_l + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, m \end{cases} \quad (24)$$

where C is the penalty coefficient. By introducing Lagrange coefficients a_i, a_i^* to equation (24), we get:

$$\begin{cases} \max w(a_i, a_i^*) = -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (a_i^* - a_i)(a_j^* - a_j)(z_i^T z_j) \\ \quad - \sigma_l \sum_{i=1}^m (a_i^* + a_i) + \sum_{i=1}^m y_i (a_i^* - a_i) \\ s.t. \sum_{i=1}^m (a_i^* - a_i) = 0, 0 \leq a_i, a_i^* \leq C, i = 1, 2, \dots, m. \end{cases} \quad (25)$$

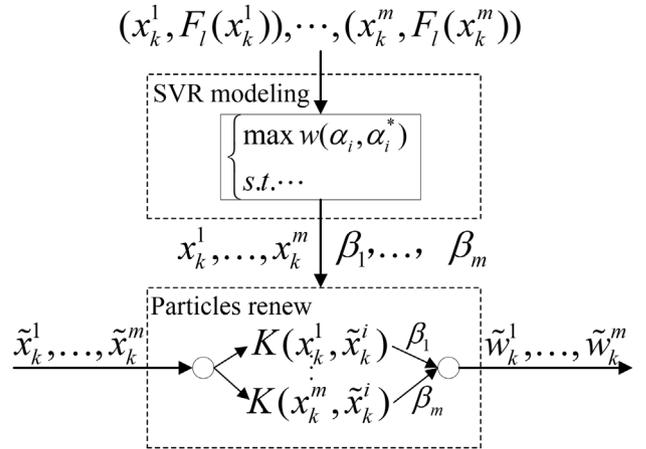


Fig. 1. Fundamental illustration of SVR-PF.

Now the solution of equation (25) can be described as:

$$\beta_j = \sum_{i=1}^m (a_i^* - a_i) z_i(x_j). \quad (26)$$

In equation (26), x_i is the support vector and the corresponding parameter of non-zero coefficients a_i^*, a_i . Substituting equation (26) into (21), the solution can be transformed into a posterior distribution estimation of an optimization problem.

As discussed above, the PF algorithm can be modified into a new PF algorithm by integrating SVR, which can be described as follows.

Resampling of the posterior distribution starts once the effective sample N_{eff} below the threshold. The two training groups are particle x_k^i and corresponding weight $w_k^i = F_l(x_k^i)$. The resampling posterior distribution is rebuilt by these groups. The flow chart of the SVR-PF algorithm is shown in Figure 1.

In Figure 1, the rebuilt particles and weights are represented by $\tilde{x}_k^1, \dots, \tilde{x}_k^m$ and $\tilde{w}_k^1, \dots, \tilde{w}_k^m$.

3 Proposed prediction method

The proposed method mainly consists of three parts, feature construction, feature signal processing and RUL prediction, see Table 2 for details.

3.1 Feature construction

Feature signals are extracted from respective original vibration and temperature signal of crusher roller sleeves. Since the definition of the original signal in different stage is relatively vague, it is crucial to select a significant sensitive feature that can fully reflect the degradation of roller sleeve. The proposed method evaluates the degradation of roller sleeves by calculating the tendency degree between each feature and running time, which is defined as the Karl Pearson coefficient of the feature.

Table 2. BPA function combination based on D-S data fusion.

Step 1: Collect the vibration and temperature signal
Step 2: Calculate the absolute correlation matrix of the selected features
Step 3: Find the feature which has the highest correlation coefficient among these features.
Step 4: Remove the interference terms of the feature signals
Step 5: Calculate each BPA function based on acceleration and temperature data
Step 6: Calculate the RUL prediction based on acceleration and temperature fusion data and repeat step 5 to step 6 until a_k equals the threshold value
Step 7: Calculate the RUL prediction $\bar{L} = (k + 1) - N$

The Karl Pearson coefficient uses the rank to evaluate the tendency degree of a feature. It cannot only evaluate the nonlinear relationship but the monotonicity of the features

$$R = \frac{\sum_{i=1}^{N_s} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N_s} (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (27)$$

where x_i and y_i are the ranks of the time t_i , and the i th feature, respectively. N_s is the length of the time sequence. \bar{x} and \bar{y} are the means of x_i and y_i , respectively.

The sensitive feature is chosen from the features which has the highest tendency value, i.e., the original feature has the most obvious monotonic trends.

3.2 Feature signal processing

In the signal processing part, original vibration and temperature signals are usually formed by the superposition of the characteristic signal and the noise signal, and the random disturbance signal in the noise affects the precision of the prediction result deeply. So, it is important to process the signal for a more accurate prediction result. Signal processing contains the removal of outliers, eliminates the trend item and denoising.

In general, the detection of outliers in signal is based on the previous normal monitoring data. The least squares polynomial is established to estimate the value of the observation data at the next moment, the absolute value of the estimated value subtracts the actual data at current time and the difference is further determined whether the difference is more than a given threshold. If the difference is more than the threshold, it is considered that the observation data are outlier, otherwise they are considered normal data.

In measurement process $\hat{x}(n) = x(n-1) + \frac{1}{2}$ or signal processing, set $x(n-4)$, $x(n-3)$, $x(n-2)$, $x(n-1)$ are four consecutive data of signal $x(n)$ before time point n . The estimated value of the current time $\hat{x}(n)$ can be obtained by the linear extrapolation of the least square estimation [22–25].

$$\hat{x}(n) = x(n-1) + \frac{1}{2}x(n-2) - \frac{1}{2}x(n-4). \quad (28)$$

Calculating the absolute value of $\hat{x}(n)$ subtracts the measured value, and compares it with the threshold value δ , i.e.

$$|\hat{x}(n) - x(n)| \leq \delta \quad (29)$$

where $x(n)$ are current data, $\hat{x}(n)$ are the estimated value of the current data obtained by the linear extrapolation through the least square estimation; σ is the standard deviation of measured data residuals. If the equation (29) is established, $x(n)$ is the normal value, otherwise, it is the outlier.

The trend term in the measurement signal is the frequency component of the signal, which is larger than the sampling length of the signal. It is generally the result of a slow change of the time sequence in the measurement system. Except the working frequency of the original signal collected by the sensor, there are some random interference signals. The existence of these trends, will cause great error in the correlation analysis or power spectrum analysis in space domain, even distort the low frequency completely. If the measurement signal without removing the trend term is directly used to predict the RUL of roller sleeve, it will directly affect the forecast results, make inappropriate judgments and conclusions. So the extraction and elimination of the measurement signal trend term is an important part of tested data processing.

The original signal is $x(t)$, which can get a discrete time series $x(n)$ by uniformly-spaced sampling. The least square method is used to construct a p th-order polynomial [26–28].

$$y(t) = a_0 + a_1t + a_2t^2 + \dots + a_pt^p = \sum_{k=0}^p a_k t^k \quad (30)$$

where p is a positive integer, means the order of polynomial, and the selection of p value is based on the estimation of the signal trend. If the trend of the signal is linear, choose $p = 1$. With $y(t)$, we subtract the original signal $x(t)$ by the polynomial trend term $y(t)$, i.e.

$$\hat{y}(t) = x(t) - y(t) \quad (31)$$

where $\hat{y}(t)$ is the signal that removes the trend item.

Because of the existence of so lot of noise, the field monitoring signal may submerge in other vibration signals and random noise, which can further causes great impact on the online monitoring. In this paper, wavelet is used to

denoise the signals. The basic idea of wavelet denoising is to decompose and reconstruct the signal. Because signal and noise at different wavelet spectrum scales have different expressions, we remove the spectral components especially the dominant portions generated by noise at different scales. The wavelet spectrum preserved in this way is the wavelet spectrum of the original signal, basically. We reconstruct the original signal using the reconstruction algorithm of wavelet transform at last [29–32].

Signal $x(t)$ can be expressed as:

$$x(t) = \sum_{k=-\infty}^{+\infty} c_{j,k} \phi(t-k) + \sum_{k=-\infty}^{+\infty} \sum_{j=0}^{+\infty} d_{j,k} c_k \varphi(2^j t - k) \quad (32)$$

where $c_{j,k} = \langle x(t), \phi_{j,k}(t) \rangle$ is the scale coefficient, $d_{j,k} = \langle x(t), \varphi_{j,k}(t) \rangle$ is the wavelet coefficient.

In the multi-scale decomposition process, $x(t)$ is always progressively decomposed to two subspaces V_j and W_j from the space V_{j-1} . According to the two-scale equation, we can get the fast recursive algorithm about projection coefficient from $c_{j-1,k}$ of $x(t)$ in V_{j-1} to $c_{j,k}$ and $d_{j,k}$ of $x(t)$ in V_j and W_j .

$$c_{j,k} = \sum_{m \in \mathbb{Z}} h(m-2k) c_{j-1,m} \quad (33)$$

$$d_{j,k} = \sum_{m \in \mathbb{Z}} g(m-2k) c_{j-1,m}. \quad (34)$$

On the contrary, $c_{j-1,k}$ also can be reconstructed by $c_{j,k}$ and $d_{j,k}$, and the reconstruction formula is as follows.

$$c_{j-1,k} = \sum_m c_{j,m} h(k-2m) + \sum_m d_{j,m} g(k-2m). \quad (35)$$

3.3 RUL prediction

3.3.1 Initial state of fusion prediction

Set N as the initial time of prediction, a_N is the acceleration in the N th time point, $a_{T,N}^*$ is the acceleration prediction obtained by temperature data in the N th time point, $a_{a,N}^*$ is the acceleration prediction obtained by acceleration data in the N th time point.

The first step is calculating the initial value of $m_{1,i}(T)$ BPA function. From the central limit theorem, a large number of temperature data measurement errors obeys normal distribution, i.e.

$$T_{T,N} \sim N(\mu_{T,N}, \sigma_T^2) \quad (36)$$

where $\mu_{T,N}$ is the mean value of temperature, σ_T^2 is the variance.

From the experimental result we can see that there is a strong linear relationship between acceleration signal and temperature signal. So, the $a_{T,N}$ estimation also obeys normal distribution, i.e.

$$a_{T,N} \sim N(\mu_{a_{T,N}}, \sigma_T^2) \quad (37)$$

where

$$\mu_{a_{T,N}} = \alpha_N * \mu_{T,N} + \beta_N = a_N \quad (38)$$

$$\sigma_T^2 = \alpha_N^2 * \sigma_T^2. \quad (39)$$

Thus the initial value of the BPA function $m_{1,i}(T)$ can be described as $m_{1,N+1}(T)$

$$m_{1,N+1}(T) = \frac{1}{\sqrt{2\pi}\sigma_T} \exp \left[-\frac{(a_{T,N}^* - a_N)^2}{2\sigma_T^2} \right]. \quad (40)$$

The second step is calculating the initial value of the BPA function $m_{2,i}(a)$. From the central limit theorem, a large number of acceleration data measurement errors obeys the normal distribution, i.e.

$$a_{a,N} \sim N(\mu_{a_{a,N}}, \sigma_a^2) \quad (41)$$

where $\mu_{a_{a,N}}$ is the mean value of acceleration; σ_a^2 is the variance. If $\mu_{a_{a,N}}$ meets:

$$\mu_{a_{a,N}} = a_N. \quad (42)$$

The initial value of the BPA function $m_{2,i}(a)$ can be described as $m_{2,N+1}(a)$

$$m_{2,N+1}(a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp \left[-\frac{(a_{a,N}^* - a_N)^2}{2\sigma_a^2} \right]. \quad (43)$$

After getting the value of BPA function $m_{1,N+1}(a)$ and $m_{2,N+1}(a)$, then we calculate the value of the other BPA functions. Because there is no correlation between the two kinds of prediction methods, it's easy to know that:

$$m_{1,N+1}(a) = m_{2,N+1}(T) = 0. \quad (44)$$

According to the properties of BPA function:

$$\begin{aligned} m_{1,N+1}(T \cup a) &= 1 - m_{1,N+1}(T) - m_{1,N+1}(a) \\ &= 1 - m_{1,N+1}(T) \end{aligned} \quad (45)$$

$$\begin{aligned} m_{2,N+1}(T \cup a) &= 1 - m_{2,N+1}(T) - m_{2,N+1}(a) \\ &= 1 - m_{2,N+1}(a). \end{aligned} \quad (46)$$

After determining all the values of the BPA function, the third step is calculating the posterior fusion BPA function. The fusion BPA function can be described as m , from formulas (6) and (7), it's easy to know that:

$$\begin{aligned} m_{N+1}(T) &= \frac{1}{1 - K_{N+1}} \sum_{B \cap C = T} m_{1,N+1}(B) m_{2,N+1}(C) \\ &= \frac{m_{1,N+1}(T) m_{2,N+1}(T \cup a)}{1 - K_{N+1}} \end{aligned} \quad (47)$$

$$m_{N+1}(a) = \frac{1}{1 - K_{N+1}} \sum_{B \cap C = a} m_{1,N+1}(B) m_{2,N+1}(C) \quad (48)$$

$$= \frac{m_{1,N+1}(T \cup a) m_{2,N+1}(a)}{1 - K_{N+1}}$$

where

$$K_{N+1} = \sum_{B \cap C = \emptyset} m_{1,N+1}(B) m_{2,N+1}(C)$$

$$= m_{1,N+1}(a) m_{2,N+1}(T) + m_{1,N+1}(T) m_{2,N+1}(a)$$

$$= m_{1,N+1}(T) m_{2,N+1}(a). \quad (49)$$

With the fusion posterior BPA function, the acceleration can be predicted in the $N+1$ th time point. Set $a_{T,N+1}^*$ is the acceleration prediction obtained by temperature data in the $N+1$ th time point, $a_{a,N+1}^*$ is the acceleration prediction obtained by acceleration data in the $N+1$ th time point, a_{N+1}^* is the acceleration prediction obtained by data fusion method in the $N+1$ th time point. Based on formulas (47)–(49), a_{N+1}^* meets:

$$a_{N+1}^* = m_{N+1}(T) a_{T,N+1}^* + m_{N+1}(a) a_{a,N+1}^*. \quad (50)$$

Thus we get the initial state of RUL prediction based on D-S data fusion and SVR-PF:

$$\text{Initial} = [m_{1,N+1}(T), m_{1,N+1}(a), m_{1,N+1}(T \cup a),$$

$$m_{2,N+1}(T), m_{2,N+1}(a) \cdots m_{2,N+1}(T \cup a),$$

$$m_{N+1}(T), m_{N+1}(a), a_{N+1}^*]^T. \quad (51)$$

3.3.2 Fusion prediction process

Step 1: calculating the BPA functions. Set N_{EOL} is the end of roller sleeve working life, a_k^* is the acceleration prediction obtained by data fusion in k th time point, $a_{T,k}^*$ is the acceleration prediction obtained by the temperature data, $a_{a,k}^*$ is the acceleration prediction obtained by the acceleration data. Where $N+1 \leq k \leq N_{EOL}$, the BPA function can be described as $m_{1,k+1}(T)$ and $m_{2,k+1}(a)$.

$$m_{1,k+1}(T) = \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left[-\frac{(a_{T,k}^* - a_k^*)^2}{\sigma_T^2}\right] \quad (52)$$

$$m_{2,k+1}(a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left[-\frac{(a_{a,k}^* - a_k^*)^2}{2\sigma_a^*}\right]. \quad (53)$$

Because there is no correlation between the two prediction models, it's easy to know that:

$$m_{1,k+1}(a) = m_{2,k+1}(T) = 0. \quad (54)$$

Based on the properties of BPA functions, there are:

$$m_{1,k+1}(T \cup a) = 1 - m_{1,k+1}(T) \quad (55)$$

$$m_{2,k+1}(T \cup a) = 1 - m_{2,k+1}(a). \quad (56)$$

Step 2: calculating the acceleration prediction in the $k+1$ th time point. From the formulas (6) and (7), formulas (52)–(56), the posterior fusion BPA function $m_{k+1}(T)$ and $m_{k+1}(a)$ can be described as:

$$m_{k+1}(T) = \frac{1}{1 - K_{k+1}} \sum_{B \cap C = T} m_{1,k+1}(B) m_{2,k+1}(C)$$

$$= \frac{m_{1,k+1}(T) m_{2,k+1}(T \cup a)}{1 - k_{k+1}} \quad (57)$$

$$m_{k+1}(a) = \frac{1}{1 - K_{k+1}} \sum_{B \cap C = a} m_{1,k+1}(B) m_{2,k+1}(C)$$

$$= \frac{m_{1,k+1}(T \cup a) m_{2,k+1}(a)}{1 - k_{k+1}}. \quad (58)$$

Thus we obtain the acceleration prediction a_{k+1}^* in the $K+1$ th time point through D-S data fusion.

$$a_{k+1}^* = m_{k+1}(a) a_{T,k+1}^* + m_{k+1}(T) a_{a,k+1}^*. \quad (59)$$

Step 3: determining whether the acceleration reaches the threshold. If not, then go back to step 1 and continue the prediction; otherwise, calculating the RUL prediction \bar{L}^* :

$$\bar{L}^* = N_{EOL} - N = (k+1) - N. \quad (60)$$

In summary, the proposed RUL prediction method based on D-S data fusion and SVR-PF can be described as flow chart in Figure 2.

A prediction model based on D-S data fusion and SVR-PF is established as:

$$\begin{cases} X_{k+1} = f(X_k, V_k) \\ Y_k = g(X_k, N_k) \end{cases} \quad (61)$$

where X_k is the prediction state, X_k, V_k are noise.

$$X_k = [X_{T,k}^T, X_{a,k}^T, X_{DS,k}^T]^T \quad (62)$$

where $X_{T,k}^T$ is the state obtained by the analysis of temperature data; $X_{a,k}^T$ is the state obtained by the analysis of acceleration data; $X_{DS,k}^T$ is the state obtained by the analysis of D-S data fusion, and there are:

$$X_{T,k} = [\lambda_{T,k}^*, T_{T,k}^*, a_{T,k}^*]^T \quad (63)$$

$$X_{a,k} = [\lambda_{T,k}^*, T_{T,k}^*, a_{a,k}^*]^T \quad (64)$$

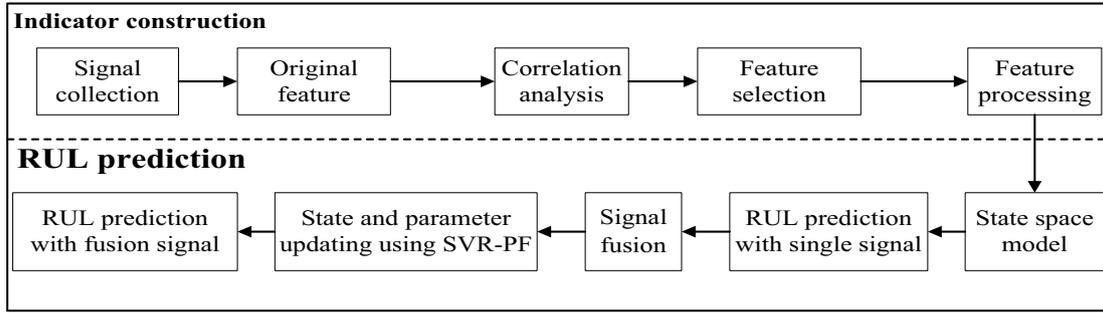


Fig. 2. Flowchart of the proposed method.

$$X_{DS,k} = [m_{1,k}(T), m_{1,k}(T \cup a), m_{2,k}(a), m_{2,k}(T \cup a) \dots m_k(T), m_k(a), a_k^*]^T. \quad (65)$$

In formulas (63)–(65): $\lambda_{T,k}$ is the degradation parameter of temperature in the k th time point [33]; $T_{T,k}$, $\lambda_{T_{T,k}}$ are the temperature degradation parameters in the k th time point [34]; * is the prediction of each corresponding variable.

Thus, in combination with the RUL prediction model of literature [33,34] we can obtain the state equation of prediction model (61), the partial state equation of prediction by temperature data can be expressed as follows.

$$\begin{cases} \lambda_{T_{T,k+1}}^* = \lambda_{T_{T,k}}^* + v_{a,k} \\ T_{T,k+1}^* = T_{T,k}^* \exp(\lambda_{T_{T,k}}^* \Delta k) + v_{b,k} \\ a_{T,k+1}^* = \alpha_N * T_{T,k} + \beta_N + v_{c,k} \end{cases} \quad (66)$$

where α_N, β_N is the degradation parameter prediction of acceleration in the N th time point through the analysis of the temperature data. The partial state equation of prediction by acceleration data can be described as follows.

$$\begin{cases} \lambda_{T_{T,k}}^* = \lambda_{T_{T,k}}^* + v_{1,k}^* \\ T_{T,k+1}^* = T_{T,k}^* \exp(\lambda_{T_{T,k}}^*) + v_{2,k}^* \\ a_{a,k+1}^* = T_{T,k+1}^* \exp(\lambda_{T_{T,k+1}}^*) + v_{3,k}^* \end{cases} \quad (67)$$

The partial state equation of prediction by data fusion can be described as follows.

$$\begin{cases} m_{1,k+1}(T) = \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left[-\frac{(a_{T,k}^* - a_k^{*2})}{2\sigma_T^2}\right] \\ m_{2,k+1}(a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left[-\frac{(a_{a,k}^* - a_k^{*2})}{2\sigma_a^2}\right] \\ m_{1,k+1}(T \cup a) = 1 - m_{1,k+1}(T) \\ m_{2,k+1}(T \cup a) = 1 - m_{2,k+1}(a) \\ K_{k+1} = m_{1,k+1}(T)m_{2,k+1}(a) \\ m_{k+1}(T) = \frac{m_{1,k+1}(T)m_{2,k+1}(T \cup a)}{1 - K_{k+1}} \\ m_{k+1}(a) = \frac{m_{1,k+1}(T \cup a)m_{2,k+1}(a)}{1 - K_{k+1}} \\ a_{k+1}^* = m_{k+1}(T)a_{T,k+1}^* + m_{k+1}(a)a_{a,k+1}^* \end{cases} \quad (68)$$

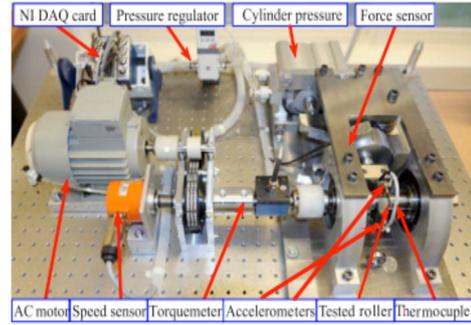


Fig. 3. Overview of the data acquisition platform.

Combining formulas (66)–(68) can obtain the state equation of prediction model (61), the measurement equation (61) can be described as follows.

$$\begin{cases} \hat{T}_{T,k}^* = T_{T,k}^* + n_{a,k} \\ T_{a,k} = T_{T,k}^* + n_{1,k}^* \\ a_k^* = m_k(T)a_{T,k}^* + m_k(a)a_{a,k}^* \end{cases} \quad (69)$$

where $\hat{T}_{T,k}$ is the temperature prediction obtained by temperature data, and $T_{a,k}$ is the degradation parameter prediction of acceleration obtained by temperature data.

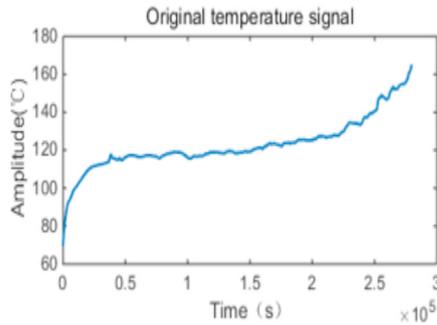
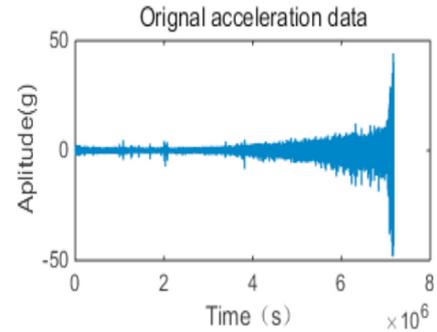
4 Experimental demonstration

4.1 Introduction to the data acquisition platform

The test platform layout is shown in Figure 3, named PRONOSTIA [35]. The testing platform is designed by the AS2M Department of the FEMTO-ST Association, the full life test of the roller sleeve is carried out on the data acquisition platform of the rolling bearing, the vibration signal is collected by the 3035B Dytran acceleration sensor (the maximum acquisition range is 50 g), temperature signal acquisition using JCJ100TLB temperature sensor (maximum acquisition range is 200 °C). Because the acceleration signal is more severe than the temperature signal, so the full-life test process stops if the acceleration signal amplitude is found to exceed 20 g. Even if the roller sleeve does not out of work, in order to avoid the test platform damage caused by the roller sleeve, we determine the failure of the roller sleeve, stop testing. The acceleration test sampling frequency is 25.6 kHz, each 10 s stores a set of

Table 3. Information about 4 groups of experimental failure roller sleeves.

Testing roller sleeve	Load (kN)	Speed (rpm/min)	Failure site	Test duration (s)
Roller sleeve 1	4	1800	The inner ring	27,560
Roller sleeve 2	4	1800	The roller	5730
Roller sleeve 3	4.2	1650	The roller	23,110
Roller sleeve 4	4.2	1650	The outer ring and roller	7530

**Fig. 4.** Original temperature data.**Fig. 5.** Original acceleration data.

data, each group of data 2560 points, the temperature test sampling frequency is 10 Hz, and each 10 s stores a set of data, each group of data 100 points.

In the test, the testing roller sleeve is 22,324 tapered roller bearing, the roller life-test is carried out 4 times, each time one test roll is damaged. The 1st test and the 2nd test are worked under the condition of radial load 4000 kN, speed 1800 rpm/min; the 3rd test and the 4th test are worked under the condition of radial load 4200 kN, speed 1650 rpm/min. The test results are shown in Table 3.

Due to the different working state of the 4 roller sleeves and the different structure of the roller sleeves, the experimental results are different, which is conform to the actual engineering facts. Then we use the measured experimental data to verify the performance of the proposed RUL prediction method. Because the 4 groups tests are carried out on the same platform, and the analysis methods of each roller sleeve is same, below takes the 1st roller sleeve as the research object to make explanation. The measured temperature and vibration data are shown in Figures 4 and 5.

4.2 Feature construction

In order to compare the prediction performance of proposed D-S data fusion and SVR-PF prediction method with the prediction method using the single acceleration data and the prediction method with the temperature data with finite data available, the first key step is to select a good feature signal.

In this paper, the features are selected by calculating the Karl Pearson correlation coefficient between the time-domain characteristics of temperature and acceleration and RUL. As an indicator the features whose correlation coefficient is highest are selected. As a result, the root mean

square (RMS) feature of vibration and the absolute mean value feature of temperature are selected. The Karl Pearson correlation coefficient result is shown in Table 4. The feature signal of acceleration and temperature are shown in Figures 6 and 7.

4.3 Feature signal processing

4.3.1 Removal of outliers

The first step of signal processing is the removal of outliers. For the nonlinear and non-stationary signal, the existence of outliers can produce spurious harmonic components, further can influence the prediction accuracy. According to the statistical properties of the original data, 3σ criterion is used to remove the outliers here. If the residuals in equation (29) exceed 3σ the outliers can be eliminated. The temperature and vibration signals after removing the outliers are shown in Figures 8 and 9.

4.3.2 Remove the trend of a smooth

Because of the zero drift of the amplifier caused by temperature variation, the performance of low frequency which exceeds the frequency range of the sensor is not stable with ambient interference around the sensor etc., which caused the collected data of vibration signal and temperature signal in life-test will often deviate from the baseline, and even the degree of the deviation from the baseline will vary over time. The whole process of the deviation from the baseline directly affects the correctness of the signal and should be removed as the trend term. This paper from the perspective of engineering application adopts a simple and practical method to remove the trend items – the modified function method. The temperature and vibration signals after removing the trend term are shown in Figures 10 and 11.

Table 4. BPA function combination based on D-S data fusion.

No.	Feature	Correlation coefficient of vibration	Correlation coefficient of temperature
1	Maximum	-0.6363	-0.6183
2	Minimum	-0.584	-0.5803
3	Absolute mean value	-0.7329	-0.7169
4	Peak	-0.613	-0.5767
5	RMS	-0.7342	-0.6832
6	Mean	-0.0224	-0.0197
7	Standard deviation	-0.7341	-0.6589
8	Skewness	-0.0631	-0.3237
9	Kurtosis	-0.0511	-0.3157
10	Variance	-0.4609	-0.3858
11	Shape factor	-0.407	-0.6836
12	Crest factor	-0.0319	-0.3248
13	Variable coefficient	-0.0169	-0.3147
14	Skew coefficient	-0.0631	-0.3801
15	Kurtosis coefficient	-0.0511	-0.4729
16	Clearance factor	-0.5372	-0.6846
17	Impact factor	-0.0487	-0.5831
18	Energy factor	-0.1358	-0.1679

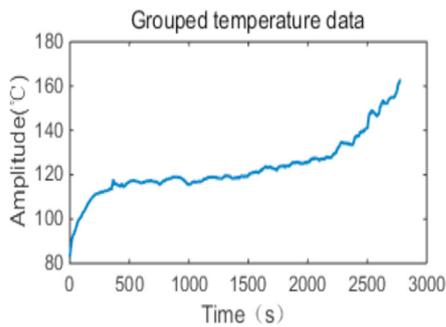


Fig. 6. Temperature feature signal.

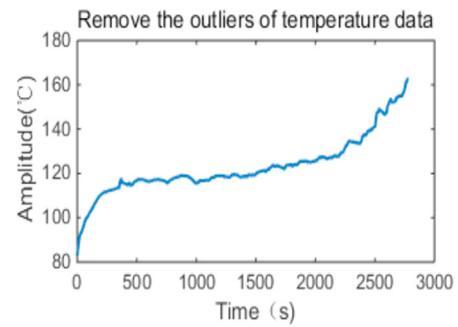


Fig. 8. Temperature signal after removing outliers.

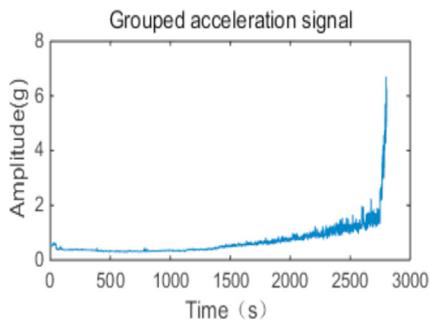


Fig. 7. Acceleration feature signal.

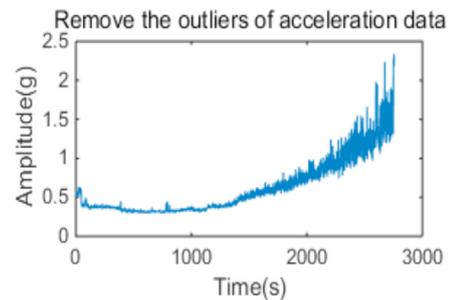


Fig. 9. Acceleration signal after removing outliers.

4.3.3 Denoising

Wavelet analysis is known as the microscope of signal processing. The key of wavelets analysis is the selection of wavelet basis and the decomposition level. Decomposition layer has great influence on the effect of denoising. The

more is the decomposition layer, the lower is the noise-signal ratio. Meanwhile, when the layer increases, the processing becomes slow. Although few decomposition layer has high noise-signal ratio, the signal is decomposed to very small frequency bandwidths. Only the high frequency coefficients can be processed to remove the

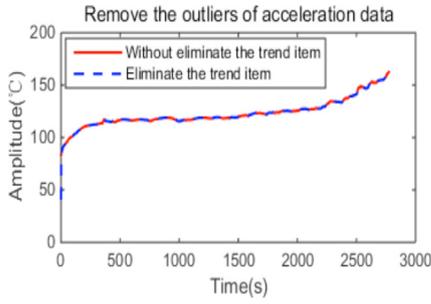


Fig. 10. Smoothed temperature signal.

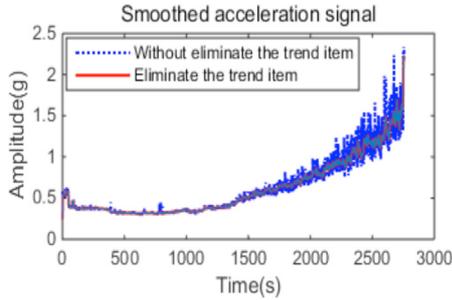


Fig. 11. Smoothed acceleration signal.

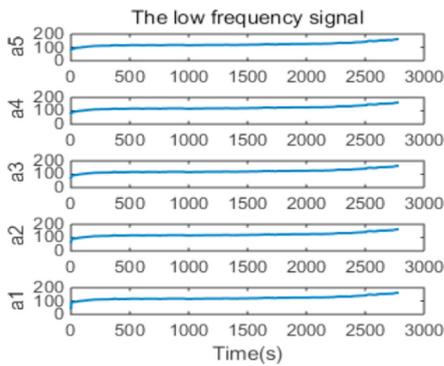


Fig. 12. Low frequency temperature signal.

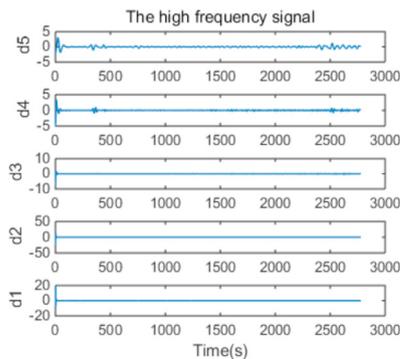


Fig. 13. High frequency temperature signal.

corresponding noise, while the corresponding low frequency noise is all reserved. Therefore, the choice of wavelet decomposition layer should be neither too large as consider-

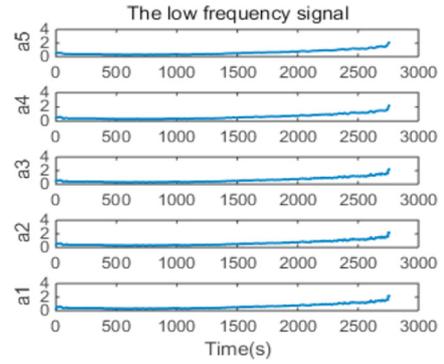


Fig. 14. Low frequency acceleration signal.

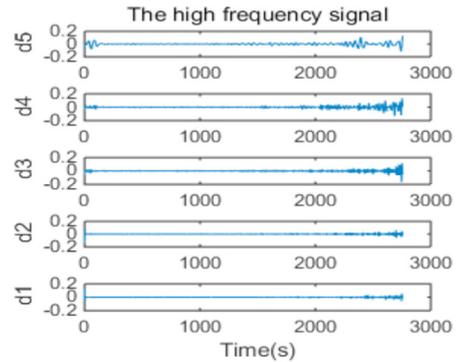


Fig. 15. High frequency acceleration signal.

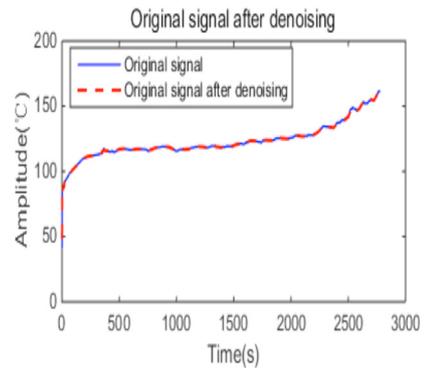


Fig. 16. Denoising temperature signal.

ing the improvement of the noise-signal ratio nor too small as considering the suppression of low frequency noise. The purpose of denoising is to get the useful feature signal, so wavelet coefficients can reflect the minimum frequency components in the useful signal. The wavelet decomposition is to decompose the signal into various independent bands, high detail coefficient reflects the low-frequency part of the signal. So this paper is based on the minimum frequency signal to determine the maximum level of wavelet decomposition. In this paper, sym8 wavelet is chosen as the wavelet base and soft thresholds are used to denoise. The temperature and vibration signal denoising processes are shown in Figures 12–17.

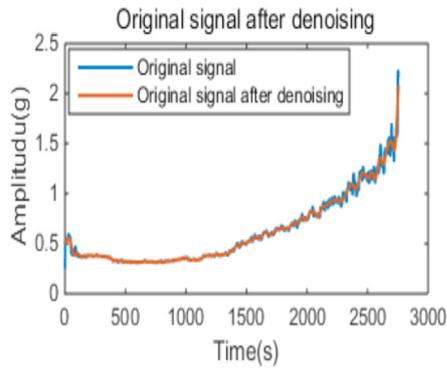


Fig. 17. Denoising acceleration signal.

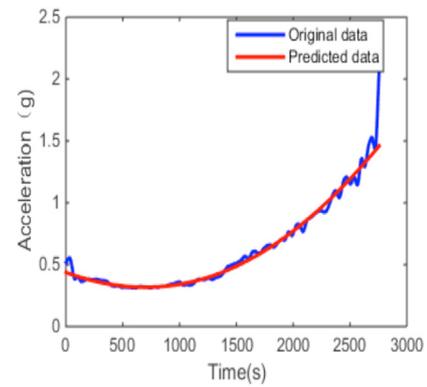


Fig. 19. Predicted acceleration by acceleration.

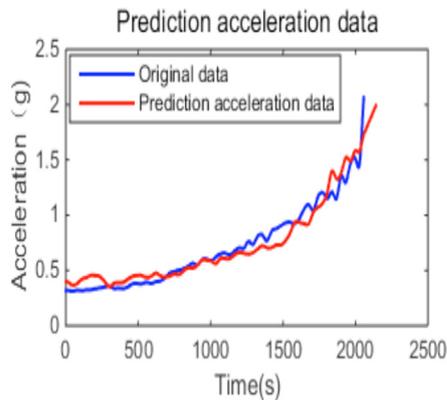


Fig. 18. Predicted acceleration by temperature.

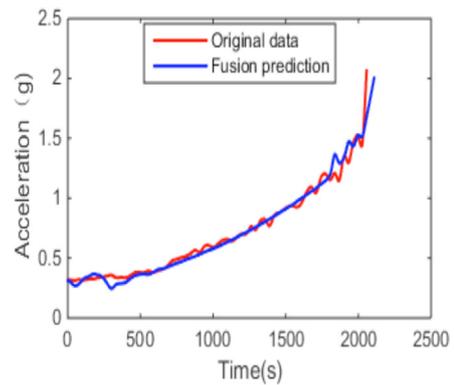


Fig. 20. Fusion prediction.

Table 5. Comparison of results of residual life prediction.

Failure roller sleeve	Roller sleeve 1	Relative error	Roller sleeve 2	Relative error	Roller sleeve 3	Relative error	Roller sleeve 4	Relative error
Tested life	27,560	0	5730	0	23,110	0	7530	0
L ₁₀ method [36]	53,250	-159%	15,313	-162%	60,086	-160%	19,392	-158%
Wang's method	1645	92%	490	91%	16,177	93%	410	95%
Edwin's method	10,691	48%	3604	37%	11,324	51%	2711	64%
RUL prediction obtained by temperature	28,420	3.12%	5984	4.43%	24,304	5.17%	7924	5.23%
RUL prediction obtained by acceleration	32,970	19.63%	6811	18.86%	28,451	23.11%	9248	22.82%
Proposed method	28,080	1.89%	5857	2.21%	23,611	2.17%	7788	3.43%

4.4 RUL prediction

We compare the prediction performance of proposed prediction method which based on D-S data fusion and SVR-PF with prediction method which uses single data source and other prediction methods. Roller sleeve 1 was used to predict in three cases, predicted by temperature data, predicted by acceleration data and predicted by fused data.

From Figures 18–20, it's easy to know that the results predicted by the proposed prediction method are more accurate than other prediction methods.

5 Conclusion

In view of the engineering problem that is difficult to accurately predict the remaining useful life under partially observed state, a new method based on D-S data fusion and SVR-PF is proposed.

From Table 5, we can see that compared to other prediction methods such as obtained by the temperature or acceleration-based data-driven, the prediction accuracy of proposed method is significantly improved. Meanwhile, it provides a basis to make maintenance

decision for the equipment working under severe conditions, further reduces the maintenance cost, improves the utilization rate and the reliability of the equipment, which has good practicability and popularization value.

Nomenclature

SOH	State of health
PHM	Prognostics and health management
BPA	Basic probability assignment
$m_1(\cdot)$	BPA function 1 under the identification framework
$m_2(\cdot)$	BPA function 2 under the identification framework
$m_{1,2}(\cdot)$	BPA function after fusion m_1 and m_2 under the identification framework
K	Degree of conflict between the two evidences
T	Temperature magnitude
a	Acceleration magnitude
a_N	Acceleration magnitude at time N
$a_{T,N}^*$	Prediction acceleration magnitude at time N obtained by temperature data
$a_{T,N}$	Acceleration magnitude at time N obtained by temperature data
$a_{a,N}^*$	Prediction acceleration magnitude at time N obtained by acceleration data
R	Value of Karl Pearson coefficient
α_N, β_N	Acceleration degradation parameters
$m_{1,i}(T)$	BPA function of temperature at time i
$a_{T,N+1}^*$	Prediction acceleration magnitude at time $N+1$ obtained by temperature data
K_{N+1}	Degree of conflict between the two evidences at time $N+1$
N_{EOL}	Threshold of life cycle N
\hat{L}^*	Prediction of remaining useful life
X_k	State of prediction at time k
V_k, N_k	Measurement noise at time k
$f(\cdot), g(\cdot)$	Transition function and measurement function
$X_{T,k}^T$	State of temperature at time k
$X_{a,k}^T$	State of acceleration at time k
$X_{DS,k}^T$	State of fusion at time k
$\lambda_{T,k}$	Temperature degradation parameters obtained by temperature at time k
$\lambda_{T,k}^*$	Prediction temperature degradation parameters obtained by temperature at time k
$T_{T,k}^*$	Prediction temperature magnitude obtained by temperature at time k
$a_{T,k}^*$	Prediction acceleration magnitude obtained by temperature at time k
$X_{a,k}$	State of acceleration at time k
$\lambda_{T_1,k}^*, T_{1,k}^*$	Acceleration degradation parameters obtained by acceleration at time k

$\lambda_{T,k+1}^*$	Prediction temperature degradation parameters obtained by acceleration at time $k+1$
$v_{1,k}$	λ degradation parameters obtained by temperature at time k
$\lambda_{T,k}^*$	Distribution state of λ at time k
$v_{2,k}$	State noise of prediction temperature obtained by temperature at time k
$v_{3,k}$	State noise of prediction acceleration obtained by temperature at time k
$v_{1,k}^*$	λ degradation parameters obtained by acceleration at time k
$v_{2,k}^*$	State noise of prediction acceleration obtained by temperature at time k
$v_{3,k}^*$	Measurement noise of prediction acceleration obtained by acceleration at time k
$n_{1,k}$	Measurement noise of prediction temperature at time k
$n_{1,k}^*$	Measurement noise of prediction acceleration degradation parameters at time k

References

- [1] X. Zhang, X. Chen, B. Li, Life prediction of machinery major equipment: a review, *J. Mech. Eng.* 47 (2011) 100–116
- [2] N.M. Vichare, M.G. Pecht, Prognostics and health management of electronics, *IEEE Trans. Compon. Pack. Technol.* 29 (2006) 222–229
- [3] Y. Lei, Z. He, Z. Yanyang, Fault diagnosis based on the new model of hybrid intelligence, *Mech. Eng.* 44 (2008) 112–117
- [4] N. Vapnik Vladimir, *Statistical learning theory*, Wiley, NY, 1998, pp. 760–768
- [5] M. Sunghwan, L. Jumin, H. Ingoo, Hybrid genetic algorithms and support vector machines for bankruptcy prediction, *Exp. Syst. Appl.* 31 (2006) 652–660
- [6] M. Nizam, M. Azah, H. Aini, Dynamic voltage collapse prediction in power systems using support vector regression, *Exp. Syst. Appl.* 37 (2020) 3730–3736
- [7] H. Dong, X. Jin, Y. Lou, Lithiumion battery state of health monitoring and remaining useful life prediction based on support vector Regression-Particle filter, *J. Power Source* 2014 (2014) 114–123
- [8] J. Llinas, D.L. Hall, An introduction to multi-sensor data fusion, *IEEE Inter. Sym. Circ. Syst.* 6 (1998) 537–540
- [9] S. Wu, W. Jiang, Research on data fusion fault diagnosis method based on D-S evidence theory, *IEEE Comp. Soc.* 1 (2009) 689–692
- [10] J. Tian, W. Zhao, R. Du, *D-S Evidence Theory and its Data Fusion Application in Intrusion Detection*, Springer, Berlin, Heidelberg, CA, 2000, pp. 244–251
- [11] H. Sorenson, D. Alspach, Recursive Bayesian estimation using Gaussian sums, *J. Auto.* 7 (1971) 465–479
- [12] B. Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman filter-particle filters for tracking applications, *IEEE. Trans. Aero. Electr. Syst.* 19 (2004) 37–38
- [13] J. Carpenter, P. Clifford, P. Fearnhead, Improved particle filter for nonlinear problems, *IEEE Proc. Radar. Sonar. Navig.* 146 (1999) 2–7
- [14] A.F. Seila, Simulation and the Monte Carlo method, *Tech.* 24 (2012) 167–168

- [15] N. Metropolis, S. Ulam, The Monte Carlo method, *J. Am. Statist. Assoc.* 60 (1948) 115–129
- [16] J. Carpenter, P. Clifford, P. Fearnhead, Improved particle filter for nonlinear problems, *IEEE Proc. Radar. Sonar Navig.* 146 (1999) 1–7
- [17] R. Kalman, A new approach linear filtering prediction problems, *Trans. ASME J. Basic Eng.* 81 (1960) 35–45
- [18] Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, CA, 2006, pp. 339–344
- [19] Z. Yinliang, Z. Changpeng, H. Bo, Z. Qinghua, Runtime support for type-safe and context-based behavior adaptation, in: Presented at Computer and Information Technology (CIT), 2012 IEEE 12th International Conference, 2012
- [20] N. Kabaoglu, Target tracking using particle filters with support vector regression, *IEEE Trans. Veh. Technol.* 58 (2009) 2569–2573
- [21] V. Vapnik, *The Nature of Statistical Learning*, Springer, Berlin, CA, 1995, pp. 225–259
- [22] G. Yao, K. Qingci, A. Yuhua, Method for eliminating data outliers based on wavelet transform, *J. Air. Spac. TT&C. Technol.* 25 (2006) 64–67
- [23] W. Lin, W. Liu, Establishment and application of spring maize yield to evapotranspiration boundary function in the Loess Plateau of China, *Agric. Waste Manag.* 178 (2016) 345–349
- [24] C. Anagnostopoulos, Quality-optimized predictive analytics, *Appl. Intel.* 45 (2016) 1–13
- [25] F. Zheng, L.Y. Liu, X.X. Liu, Y. Li, X.G. Shi, G.Y. Zhang, K. W. Huan, Study on outliers influence in NIR quantitative analysis model, *Guang Pu Xue Yu Guang Pu Fen Xi* 36 (2016) 3523–3529
- [26] P. Zhang, J. Chang, B. Qu, Q. Zhao, Denoising and trend terms elimination algorithm of accelerometer signals, *Math. Prob. Eng.* 2016 (2016) 1–9
- [27] E.L. Andreas, G. Treviño, Using wavelets to detect trends, *J. Atmos. Ocean Technol.* 12 (1997) 554–564
- [28] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, *Int. J. Control* 50 (1989) 1873–1896
- [29] C. Torrence, G.P. Compo, A practical guide to wavelet analysis, *Bull. Am. Meteor. Soc.* 79 (1998) 61–78
- [30] D.E. Newland, Wavelet analysis of vibration: Part 1—Theory, *J. Vib. Acoust.* 116 (1994) 21–37
- [31] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Theory* 36 (1990) 961–1005
- [32] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, *Mech. Syst. Sig. Process* 18 (2004) 199–221
- [33] D. Hancheng, J. Xiaoning, L. Yangbing, Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter, *J. Powder Source* 2014 (2014) 114–123
- [34] W. Changhong, D. Hancheng, Remaining effective working time prediction method for vehicle lithium ion battery, *J. Auto. Eng.* 37 (2015) 476–479
- [35] Y. Lei, A model-based method for remaining useful life prediction of machinery, *IEEE Trans. Relia.* 65 (2016) 1–13
- [36] Y. Jie, Z. Xiaodong, Analysis and application of rolling bearing life calculation method, *Petro. Mach.* 32 (2004) 27–29

Cite this article as: H. Liu, J. Wu, X. Ye, T. Liao, M. Chen, A method based on Dempster-Shafer theory and support vector regression-particle filter for remaining useful life prediction of crusher roller sleeve, *Mechanics & Industry* **20**, 106 (2019)