Engineering Applications of Artificial Intelligence 116 (2022) 105391

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Reliability assessment and prediction of rolling bearings based on hybrid noise reduction and BOA-MKRVM

Shuzhi Gao [a], Yifan Yu [b], Yimin Zhang [a,*]

[a] Equipment Reliability Institute, Shenyang University of Chemical Technology, Shenyang 110142, PR China
[b] College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, PR China

ARTICLE INFO

Keywords:
Intrinsic time-scale decomposition
Maximum correlation kurtosis deconvolution
Bayesian optimization
Relevance vector machine
Logistic regression
Rolling bearing reliability model

ABSTRACT

To figure out the problem of reliability assessment and prediction due to the noise in rolling bearing vibration signals, bearing reliability assessment and prediction model was proposed combine intrinsic time-scale decomposition-adaptive maximum correlation kurtosis deconvolution (ITD-AMCKD) and Bayesian optimization algorithm mixed kernel relevance vector machine(BOA-MKRVM). First, the ITD-AMCKD hybrid model is come up with to decrease noise and extract valid information of bearing vibration signal; secondly, set up the reliability assessment model of bearing through logistic regression model; in the end, use BOA-MKRVM model to predict bearing degraded state, then substitute consequence into established reliability assessment model to acquire the prediction consequence of the bearing's reliability. The test data from Xi'an Jiaotong University demonstrate availability of model in the paper.

1. Introduction

Denoising Super Utile!

Rolling bearings apply in many aspects such as national economy and defense industry. They are essential basic components of rotating machinery and it will impact the regular operation of rotating machinery straightly. Therefore, it is crucial to analyze reliability of bearings rightly. Vibration signals of bearings can show degradation state and tendency of bearings timely (Wang et al., 2018; Attoui et al., 2020; Sun and Cao, 2020), which can be used as basics for assessment and prediction of reliability.

Bearing vibration signals include lots of significance information of bearings. However, in engineering practice, the operation of large machinery is often together with lots of noise, so the removal of rolling bearing noise and vibration signals is significant in extracting effective information of bearings. Traditional denoising algorithms, such as wavelet transform (WT) (Bao et al., 2020; Wang et al., 2021), wavelet packet transform (Li et al., 2021; Malhotra et al., 2021), empirical mode decomposition (EMD) and its deformation (Fan et al., 2020; Liu et al., 2019; Hou et al., 2020), local mean decomposition (LMD) (Ning et al., 2016; Han et al., 2021), local characteristic scale decomposition (LCD) (Bian, 2017; Ding et al., 2020) and other time–frequency analysis methods.

Aside from time–frequency analysis means, also have many great denoising means. Minimum entropy deconvolution (MED) (He et al., 2020; Xu et al., 2019) can wipe off the noise in signal, and enhance kurtosis value and SNR. There are many valid and unique blind source separation algorithms. Include fast independent component analysis (FastICA) (Miao and Zhao, 2020a), constrained independent component analysis (CICA) (Hao et al., 2019; Li et al., 2020a) and joint approximate diagonalization of eigenmatrices (JADE) (Miao et al., 2020; Islam et al., 2018). In addition, based on a single filter, the researchers come up with hybrid filter. Bi et al. (2015) used RobustICA, continuous wavelet transform (CWT) in order to isolate noise source of gasoline engines. Using this method accelerates the speed of separating the noise source of the gasoline engine. Miao and Zhao (2020a) aimed at the difficulty of separating mixed signals with traditional signal processing methods, which brought difficulties to machine health monitoring and fault diagnosis, and proposed a combination of synchronous cumulative average (SCA) noise reduction and improved FastICA.

After using the method to decrease the noise of the bearing signal, the effective signal can be got, then through the reliability model to assessment. Weibull proportional hazard model (WPHM), logistic regression model (LRM) and Monte Carlo (MC) are familiar reliability assessment models. Qiu et al. (2019) proposed an integrated life forecast model combine on improved support vector regression (SVR) and WPHM to forecast overall life of the bearing. Compared with the previous methods, the effect of using SVR and WPHM models to achieve life of bearing forecast has been further improved. Gao et al. (2020) proposed a rolling bearing operation reliability assessing and predicting method based on isometric mapping and nonhomogeneous cuckoo search-least squares support vector machine (NoCuSa-LSSVM). Far et al. (2021) come up with a probability method for predicting the spread of COVID-19 through time-varying reliability analysis and Monte Carlo.

* Corresponding author.
E-mail address: zhangyimin_126163@126.com (Y. Zhang).

https://doi.org/10.1016/j.engappai.2022.105391
Received 27 April 2022; Received in revised form 24 July 2022; Accepted 24 August 2022
Available online 10 September 2022

https://www.sciencedirect.com/science/article/abs/pii/S221478532039845X

Come già evidenziato per gli approcci Data Driven

After the reliability assessment, by predicting state of performance to forecast reliability. Current bearing prediction methods are mainly as follows: data-driven method and model-based method. Data-driven is combines experimental data with machine learning. Wang et al. (2019) work out residual life prediction method fused feature parameter extraction and long short-term memory (LSTM) neural network. Ali et al. (2015) combined WPHM and artificial neural network (ANN) to forecast life of bearings precisely. E et al. (2021) proposed mixture extreme-point symmetric mode decomposition (ESMD), kernel ICA and least squares support vector regression (LSSVR) to predict the carbon price. Mao et al. (2019) proposed a remaining useful life (RUL) prediction method based on shrinking denoising autoencoder and transfer learning to solve problem of less bearing degradation data, and unsatisfactory remaining life prediction results. The results indicate that when using this method, there is a valid improvement in prediction accuracy and numerical stability. Pan et al. (2020) proposed two-stage prediction method and apply multi-variable feedback extreme value learning machine model to predict the RUL of rolling bearings accurately. Meng et al. (2020) used differential empirical mode decomposition (DEMD) and grey Markov model-based decomposition methods to define parameters that can indicate the degradation trend of rolling bearings and improve the prediction accuracy of RUL. Yang et al. (2020) proposed new method to predict RUL of the auto regress-unscented kalman filter (AR-UKF) model and the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). Huang et al. (2019) used a new forecast method based on bidirectional long short-term memory (BLSTM) to forecast RUL of engineering systems. Li et al. (2020b) put forward data segmentation algorithm using sparse low-rank matrix (SLRM) and a chaotic bionic optimization algorithm based on bearing health index time series (HITS) for the prediction of bearing degradation trends, and combined with the SVM model predict the time series.

Tu cercale!

Model-based methods use mathematical or physical models that have been established before to represents the process of change. Wang et al. (2016) based on enhanced KF and expectation maximization algorithm establish a bearing degradation predict model well.

The above paper makes a profound study on the degradation trend of bearings, but many researchers ignore the defects of the noise reduction algorithm itself on the noise reduction effect, making the noise reduction effect is not good, many researchers have not put forward a better solution to the noise removal problem in the original vibration signal of bearings, the research process is not deep enough. According to the retrieval of mentioned papers, it can know that the noise reduction method used in the above-mentioned documents still cannot resolve the noise problem of the vibration signal well. In this study, a mixed denoising and algorithm based on intrinsic time-scale decomposition-adaptive maximum correlation kurtosis deconvolution (ITD-AMCKD) combine Bayesian optimization algorithm mixed kernel relevance vector machine (BOA-MKRVM) is proposed for bearing reliability assessment and prediction model. Firstly, the vibration signal with noise is decomposed by ITD to get modal components PRs. In this study, an evaluation index is proposed to realize the automatic selection of PRs, and the effective information of bearing vibration signal is extracted by using the improved MCKD algorithm for secondary denoising of bearing signal. Next, BOA is used to optimize the mixed kernel RVM parameters, and BOA-MKRVM prediction model is used to assess and predict bearing reliability. The major innovations and efforts is:

1. Aiming at vibration signal noise problem, put forward ITD-AMCKD hybrid method to fall off extra information of the noisy signal.
2. Aiming at selection PR ? after ITD, combines the refined composite multiscale dispersion entropy, the root mean square error of each component, and correlation coefficient, and proposes an evaluation index to realize the automatic selection of PRs, which improves noise reduction efficiency and accuracy of ITD method.
3. Aiming at the selection of displacement number M and filter order number L in MCKD, used sparrow optimization algorithm to improved MCKD, which further improves effect of MCKD method and the SNR of the processed signal obviously.
4. Aiming at the prediction of bearing reliability, proposes BOA-MKRVM model to predict the reliability of bearing signals, and improved accuracy obviously.

This paper structure is as follows: the I part is the introduction, the second section introduces ITD and AMCKD hybrid noise model, the third section introduces the logistic regression bearing reliability model, the fourth section introduced the BOA to optimize MKRVM prediction model, the fifth section of the model is verified through experiments, the sixth section of the conclusions of this paper.

2. Hybrid noise reduction model of rolling bearing

2.1. *Intrinsic Time-scale Decomposition (ITD)*

Intrinsic Time-scale Decomposition (Frei and Osorio, 2007) can decompose the signal into the amount of several inherent rotating components PRs and a trend component r. ITD algorithm can not only reduce the phenomenon of mode aliasing in signal decomposition, but also reduce the complex screening and spline interpolation, which can enhance the signal processing efficiency, also can realize the function of reflecting the frequency change in real time. Assuming X_t is the input, p is baseline extraction operator, The decomposition process is:

$$X_t = pX_t + (1-p)X_t = p_t + H_t \tag{1}$$

p_t is the baseline component; H_t is the rotation component.

The decomposed high-frequency rotation component is removed, the baseline component signal is regarded as the next signal to be decomposed, and the above decomposition process is iterated until a monotonic trend component signal appears. The whole decomposition process of X_t is:

$$X_t = pX_t + HX_t = (H\sum_{l=0}^{m-1} p^l + p^m)X_t \tag{2}$$

pX_t is intrinsic rotation component; HX_t is the fixed rotation component extraction operator; Hp^lX_t is the $l+1$ rotation component; p^mX_t is the monotonic trend component.

2.2. *Adaptive Maximum Correlation Kurtosis Deconvolution (AMCKD)*

2.2.1. *Maximum Correlation Kurtosis Deconvolution (MCKD)*

Maximum Correlation Kurtosis Deconvolution (MCKD) is an algorithm improved by Mcdonald et al. (2012) from Minimum Entropy Deconvolution. MCKD can perform blind deconvolution of signals. Compared with MED, MCKD not only considers the continuity of the signal, but also highlights the periodicity of the signal, and uses the iterative method to conduct the deconvolution operation of the signal to remove noise. Understanding MCKD first requires understanding the concept of correlation kurtosis. The correlation kurtosis is:

$$CK_M(T) = \frac{\sum_{n=1}^{N} (\prod_{m=0}^{M} y(n-mT))^2}{(\sum_{n=1}^{N} y(n)^2)^{M+1}} \tag{3}$$

T is the number of sampling points covered in a single cycle, M is the delay, N is the length of the signal, $y(n) = \sum_h^H f(h)x(n-h+1)$, when n ≠ 1, 2…N, x(n) = 0, y(n) = 0. And f is the length of the filter.

The goal of MCKD is to maximize the correlation kurtosis, and then the resulting filter length f is the optimal solution. The MCKD objective function is:

$$maxCK_M(T) = max\frac{\sum_{n=1}^{N} (\prod_{m=0}^{M} y(n-mT))^2}{(\sum_{n=1}^{N} y(n)^2)^{M+1}} \tag{4}$$

Fase 1: Denoising

1. Il segnale di vibrazione con Rumore
   viene decomposto da ITD (Intrinsic Time Scale Decomposition)

   ↓

   Vengono estratte così le componenti modali PR1

2. L'estrazione delle features avviene poi tramite l'algoritmo MCKD migliorato per un secondo denoising!

3. BOA - MKRVM per predire l'affidabilità del pezzo.

To get the optimal filter equation that maximizes $CK_M(T)$, take the derivative with respect to $CK_M(T)$:

$$\frac{d}{df_k} CK_M(T) = 0 \qquad k = 1, 2, \dots, L \tag{5}$$

→ Banale Problema di Ottimizzazione

The result obtained is:

$$f = \frac{\left\| y^2 \right\|}{2\|\beta\|^2} (X_0 X_0{}^T)^{-1} \sum_{m=0}^{M} (X_m{}^T \alpha_m) \tag{6}$$

$$X_g = \begin{bmatrix} x_{1-g} & x_{2-g} & x_{3-g} & \cdots & x_{N-g} \\ 0 & x_{1-g} & x_{2-g} & \cdots & x_{N-1-g} \\ 0 & 0 & x_{1-g} & \cdots & x_{N-2-g} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x_{N-L-g+1} \end{bmatrix}_{L \times M} \tag{7}$$

$$\alpha_m = \begin{bmatrix} y^{-1}_{1-mT_S} \left(y_1^2 y^2_{1-T_S} \cdots y^2_{1-MT_S} \right) \\ y^{-1}_{2-mT_S} \left(y_2^2 y^2_{2-T_S} \cdots y^2_{2-MT_S} \right) \\ \cdots \\ y^{-1}_{N-mT_S} \left(y_N^2 y^2_{N-T_S} \cdots y^2_{N-MT_S} \right) \end{bmatrix} \tag{8}$$

$$\beta = \begin{bmatrix} y_1 y_{1-T_S} & \cdots y_{1-MT_S} \\ y_2 y_{2-T_S} & \cdots y_{2-MT_S} \\ \cdots & \\ y_N y_{N-T_S} & \cdots y_{N-MT_S} \end{bmatrix} \tag{9}$$

$g = mT_s, m = 1, 2 \dots M.$

The MCKD process is as follows:

(1) Determine the filter length L, the delay number M, and the number of sampling points T;

(2) Find the $X_0 X_0^T$ and $X(mT)$ of the original signal;

(3) Seeking the output signal $y(k)$;

(4) Calculate α_m and β from $y(k)$;

(5) The new filter length f is obtained, and the difference between correlation kurtosis before and after the filtering is calculated. If the difference is less than the set threshold, the iteration is ended, otherwise (3) is returned.

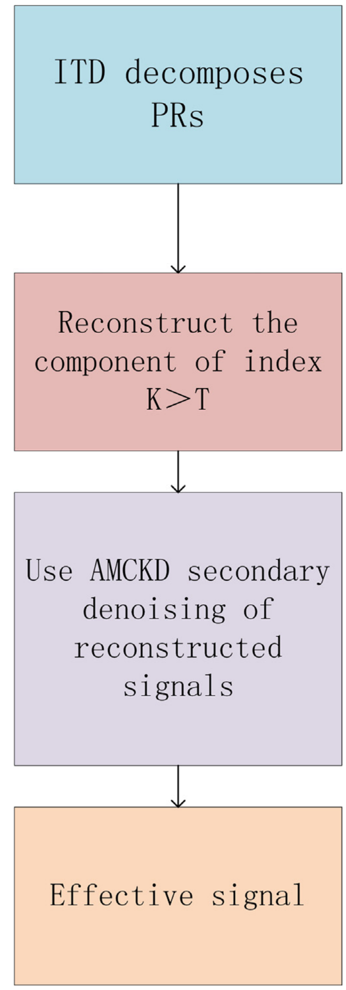


Fig. 1. Steps of hybrid noise reduction.

2.2.2. Adaptive MCKD

In MCKD, delay time M and the filter order L determine the filtering effect. To avert affect of human subjective factors on MCKD parameter setting, the sparrow search algorithm (SSA) is used to select two key influencing parameters (M and L) in MCKD.

SSA (Xue and Shen, 2020) was enlightened by sparrows foraging behavior. This algorithm is very novel and has better searching ability than other algorithms. In this algorithm, the sparrow is divided as discoverer, follower and vigilante, and their positions correspond to a solution, respectively. The discoverer found foraging directions for the entire population, and the followers followed by the foraging, and the vigilante responsible for protecting the foraging from other animals.

Assume that the sparrow population is $X = \{x_1, x_2, \dots x_n\}$. The number of sparrows is n. All sparrows fitness value of sparrows can be expressed in the following form:

$$F_x = \begin{bmatrix} f\left(\begin{bmatrix} x_1{}^1 & x_1{}^2 & \dots & x_1{}^d \end{bmatrix} \right) \\ f\left(\begin{bmatrix} x_2{}^1 & x_2{}^2 & \dots & x_2{}^d \end{bmatrix} \right) \\ \dots \\ f\left(\begin{bmatrix} x_n{}^1 & x_n{}^2 & \dots & x_n{}^d \end{bmatrix} \right) \end{bmatrix} \tag{10}$$

d is the dimension of the variables to be optimized.

The discoverer's mission is to find food and furnish foraging directions for participants. It would be farther away than his other participants flew. So, the finder's foraging search area is larger than that of the entrant. The discoverer's place change is as follows:

$$X_{i,j}{}^{t+1} = \begin{cases} X_{i,j} * exp\left(-\frac{i}{\alpha.M} \right), if \quad W < S \\ X_{i,j} + R.G, \qquad if \quad W \geq S \end{cases} \tag{11}$$

t is the number of current iterations, M is the largest number of iterations, the α between 0–1. R is the random number obeying a normal distribution, W and S are warning and safety values. When W < S, it indicates that the area when looking for food is safe and can be widely searched. G is 1 by d matrix where each entry is 1. When W ≥ S, it means that the area is no longer safe, and there may be their natural enemies, so evacuate here.

During the search for food, the follower will monitor the discoverer at all times. When they saw the discoverer find a better food, they move away from their present location and contend it. If they win, they will get these food, otherwise some hungry entrants fly elsewhere. The follower's location update is as follows:

$$X_{i,j}{}^{t+1} = \begin{cases} R.exp(\frac{X_W - X_{i,j}{}^t}{i^2}), & if \quad i > n/2 \\ X_B{}^{t+1} + \left| X_{i,j} - X_B{}^{t+1} \right| .C^{+}.G, otherwise \end{cases} \tag{12}$$

X_B and X_W represent the best and worst positions occupied by the discoverer respectively. C is 1 × d matrix, where value is allocated 1 or −1, and $C^+ = C^T(CC^T)^{-1}$. When i > n/2, it indicates that the ith participant is less fit and is at a disadvantage to compete with the discoverer for food, so that they will leave the place and go somewhere else.

The task of vigilante in the population is to protect other sparrows from their natural enemies during foraging. Its location is updated as

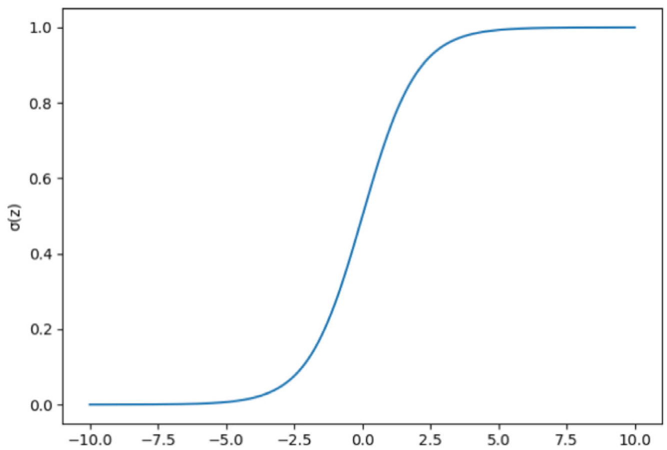


Fig. 2. Sigmoid function.

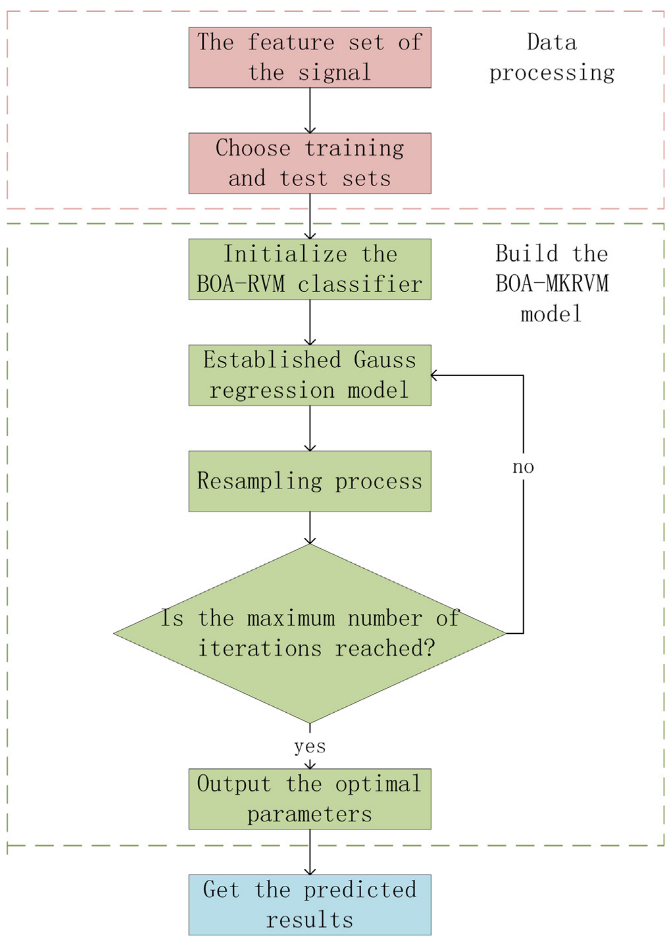


Fig. 3. BOA-MKRVM prediction model.

follows:

$$X_{i,j}{}^{t+1} = \begin{cases} X_{best}{}^{t} + \beta \left| X_{i,j}{}^{t} - X_{best}{}^{t} \right|, if \quad f_i > f_b \\ X_{i,j}{}^{t} + A(\frac{\left| X_{i,j}{}^{t} - X_{W}{}^{t} \right|}{(f_i - f_b) + \epsilon}), if \quad f_i = f_b \end{cases} \tag{13}$$

X_{best} is the best position, $\beta \sim N(0,1)$ to control step length, f_i is the fitness value, f_b and f_o are global best and worst fitness values, ϵ is a constant, to prevent the denominator of 0, A $\in [-1-1]$, When $f_i > f_b$, it was indicated that it was located at the edge of the population, in order to improve the fitness. When $f_i = f_b$, it was indicated that it was in a central position and was going close to its surrounding companions to protect them.

The SSA updates the positions of the three to find the optimal location and complete food predation.

2.3. Hybrid noise reduction model

The analysis of PRs decomposed by ITD algorithm depends on user experience, and PR needs to make artificial choices. In order to make PR analysis faster and more automatic, in our study proposes an evaluation index composed of Refined Composite Multiscale Dispersion Entropy (RCMDE), Root Mean Square Error (RMSE) of each PR component, and Pearson correlation coefficient between the original signal and PRs.

(1) RCMDE is proposed by Azami et al. (2017) on the basis of Dispersion Entropy (DE) theory (Rostaghi and Azami, 2016). Subsequently, some scholars proved that the calculation error of fine composite multi-scale dispersion entropy was smaller than that of dispersion entropy, which could better show the variation trend of vibrate signals in different states.

Because multiscale analysis can reflect the complexity characteristics, the bigger the RCMDE is, the stronger the complexity of the signal is. Therefore, calculating the RCMDE of each PR can well characterize the state information of this PR. In our study, the average RCMDE of all PR components is calculated to screen the effective PR components. RCMDE calculation steps are as follows:

Step 1: For original data u, its kth coarsely granulated array $x_k(x) = \{x_{k,1}(\tau), x_{k,2}(\tau), \ldots x_{k,L}(\tau)\}$ can be expressed as:

$$x_{k,j}(\tau) = \frac{1}{\tau} \sum_{b=k+\tau(j-1)}^{k+j\tau-1} u_b \tag{14}$$

Step 2: For different scale τ, the RCMDE is showed as:

$$E_{RCMDE}(X, m, c, d, \tau) = -\sum_{\pi=1}^{c^m} \overline{T} ln(\overline{T}) \tag{15}$$

$$T = \overline{P}(\pi_{v_0 v_1 \ldots v_{m-1}}) = \frac{1}{\tau} \sum_{k=1}^{\tau} P_k{}^{(t)} \tag{16}$$

(2) Pearson correlation coefficient can judge the degree of correlation. When P value is less than 0. 05, it is directly judged that the two variables are not correlated without considering other factors. Pearson correlation coefficient between original signal and PRs is expressed as follows:

$$Per_{X,X_i} = \frac{cov(X, X_i)}{\sigma_X \sigma_{X_i}} \tag{17}$$

σ_X and σ_{X_i} represent the MSE between the source signal and ith PR component after decomposition.

(3) RMSE is the deviation between measured value and sample value. The lower the value, the smaller the deviation. RMSE between vibration signal X and PR component x_i signal is expressed:

$$RMSE_{X,X_i} = \sqrt{\frac{\sum_{j=1}^{N} (X(j) - X_i(j))^2}{N}} \tag{18}$$

(4) According to the above three coefficients, this paper proposes a PR evaluation index K to screen out the PR components with more effective information. Because the number of PR obtained by decomposition of each signal is different, the amount of filtered components is also different. After you calculate the K value of each PR, you calculate the mean of the population K, which is called T. The component whose K value of PR component is greater than T value is selected and reconstructed as the result of ITD decomposition.

$$K = \frac{Per_{x,x_i} * \frac{\sum_{i=1}^{L-1} RCMDE_i}{L-1}}{RMSE_{x,x_i}} \qquad N = \frac{\sum_{i=1}^{L-1} K}{L-1} \tag{19}$$

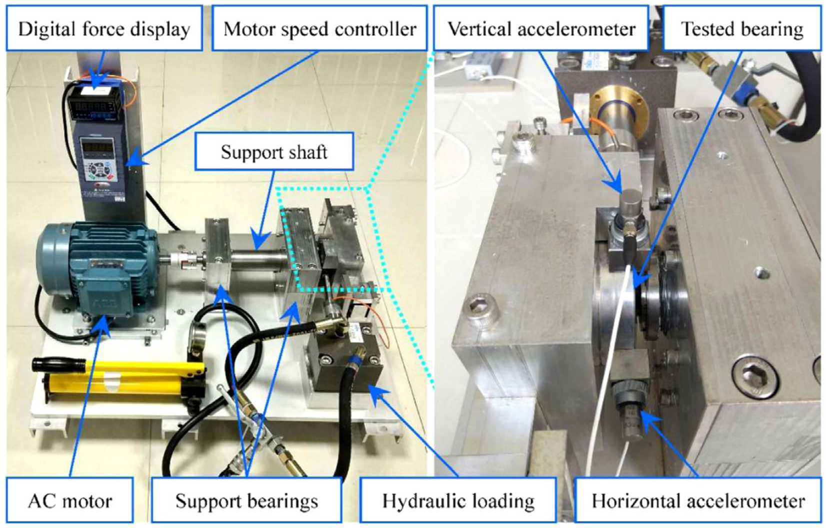
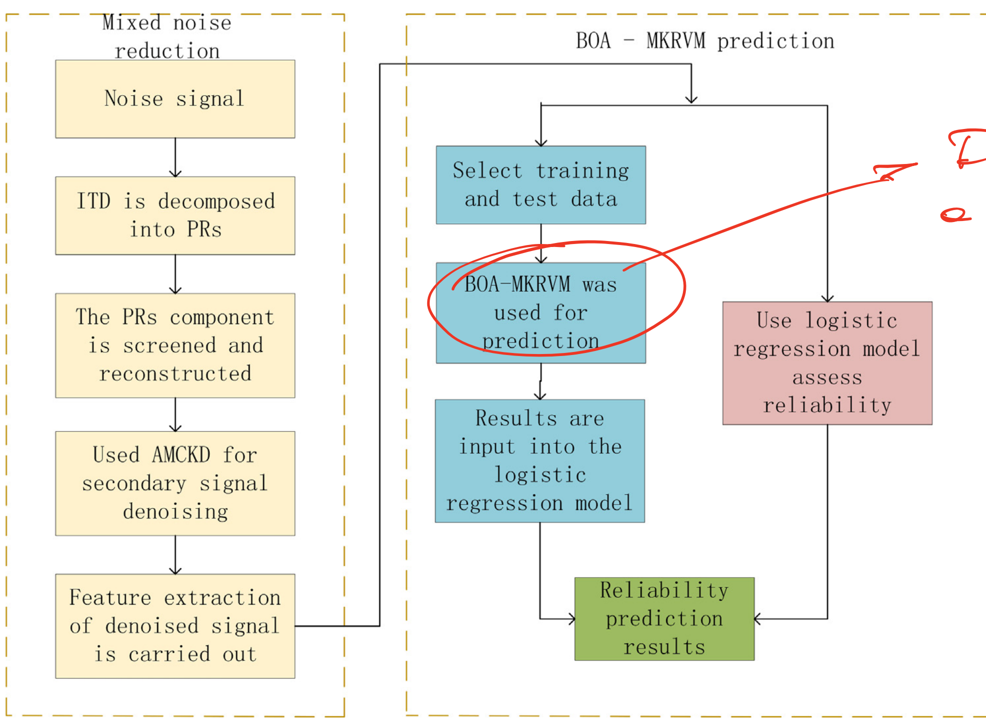


Fig. 4. Bearing test bench.



De reevaluat
a secțiunea

Fig. 5. Experimental flow.

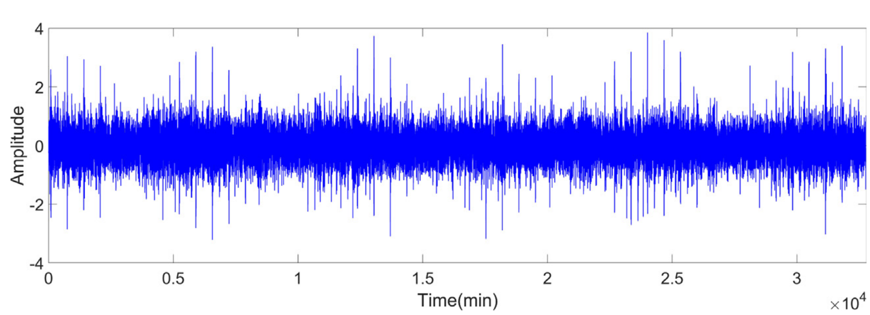


Fig. 6. Original noisy signal.

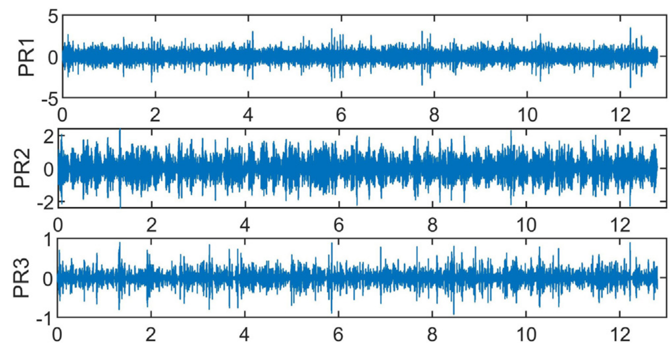


Fig. 7. PR1, PR2, PR3.

L is the amount of PR components. X represents raw signal and x_i represents the ith PR component.

To sum up, in this paper, the signal filtered by PR evaluation index K was reconstructed, and the result was used as the input signal of AM-CKD for secondary denoising. The steps of secondary noise reduction is shown in Fig. 1:

3. Rolling bearing reliability assessment model

Logistic regression model can estimate the probability through the analysis of characteristic parameters, which is a common method in mechanical equipment reliability evaluation. The logistic regression model controls the input between 0 and 1 through the sigmoid function, which is a monotonically increasing function as shown in Fig. 2. According to the definition of reliability,assuming that the ith dimension feature parameter set at time t is $X_i(t) = (x_1(t), x_2(t) \ldots x_i(t))$. Bearing runs normally $y(t) = 1$ at point t, $y(t) = 0$ when the bearing fails. Input x_i into Sigmoid function, bearing reliability is calculated:

$$R(t|X_i) = P(y_i = 1|X_i) = \frac{exp(\beta_0+\beta_1 x_1(t)+\beta_2 x_2(t)+\cdots+\beta_i x_i(t))}{1+exp(\beta_0+\beta_1 x_1(t)+\beta_2 x_2(t)+\cdots+\beta_i x_i(t))} \tag{20}$$

$\beta_0, \beta_1, \ldots \beta_i$ is regression coefficients for eigenvector set. The regression coefficients are solved using the common maximum likelihood estimation method, first transforming the above equation:

$$\ln \frac{P(y_i=1|X_i)}{1-P(y_i=1|X_i)} = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \cdots + \beta_i(t) \tag{21}$$

Put C $= \beta_0, \beta_1, \ldots \beta_i$ bring into the above formula:

$$\ln L(C) = \sum_i \left[y_i C X(t) - \ln(1 + \exp(C X(t)))\right] \tag{22}$$

Using the gradient descent method to solve the above equation, the regression coefficient can be obtained, and the reliability model can be established by bringing it into the reliability formula. The principle of the logistic regression model is simple, the regression parameters are easy to solve, and the model has strong interpretability and generalization ability. Therefore, this paper uses a logistic regression model as a reliability evaluation model.

4. Rolling bearing reliability prediction model

4.1. Bayesian Optimization Algorithm

(BoA)

Bayesian Optimization Algorithm (BOA) is a global optimization algorithm, the goal is to seek out global optimal result. By establishing alternative probability model of objective function, Bayesian optimization seeks for the hyperparameters with the best performance in the proxy model, then applies the hyperparameters into the practical objective function, and updates the proxy model on the basis of the consequence of the practical objective function, which decreases the call to the objective function and makes tuning speed and results more efficient. This is an agent-based optimization approach. We use a Gaussian process to model the proxy function. The steps of Bayesian optimization are as follows:

Step 1: Gaussian process is estimated and updated through sample point $G(x_i, y_i)$.

According to Bayesian theory, the prior function of sample set G established by Gaussian process will become a posterior distribution under the condition of given test sample is the kernel matrix of order N*1 covariance between test sample $x*$. Therefore, the joint distribution of the target value and predicted value of the sample data set can be expressed as:

$$\begin{bmatrix} y \\ f(x*) \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + {\sigma_n}^2 I & K(X,x*) \\ K(x*,X) & K(x*,x*) \end{bmatrix}\right) \tag{23}$$

$K(X,x*) = K^T(x*,X)$ is kernel matrix of order N*1 covariance between test sample $x*$ and training set X. Thus, we can get the posterior distribution that the predicted value $f(x*)$ meets in the test sample $x*$ model:

$$P(f(x*)|X,Y,x*) \sim N(\mu*, \Sigma*) \tag{24}$$

$$\mu* = K(x*,X)[K(X,X) + {\sigma_n}^2 I]^{-1} Y \tag{25}$$

$$\Sigma* = K(x*,x*) - K(x*,X)[K(X,X) + {\sigma_n}^2 I]^{-1} K(X,x*) \tag{26}$$

$\mu*$ is the mean value; $\sum *$ is the variance; I is an n-dimensional identity matrix.

In summary, $\theta = \{{\sigma_f}^2, {\sigma_n}^2, \wedge\}$ is a set of hyperparameters, and Gaussian process regression uses maximum likelihood method to determine the optimal hyperparameter combination, and its log-likelihood function is:

$$L(\theta) = -log[p(Y|X,\theta)] = \frac{1}{2} Y^T M^{-1} Y + \frac{1}{2} log|M| + \frac{n}{2} log 2\pi \tag{27}$$

M can be expressed as $K(X,X) + {\sigma_n}^2 I$

Step 2: Guide the new sampling by extracting function $a(x)$

The sampling function is the reference evidence for the Bayesian optimization method to obtain the next sample point in the hyperparameter space. Bayesian optimization, as an efficient global optimization algorithm, can randomly sample and calculate in a given hyperparameter interval, firstly obtain the preliminary distribution of the function to be optimized, and get the optimal objective function $f(x)$ through internal optimization solution $x*$. The article uses the PI function as the sampling function, the equation is:

$$f_{PI}(x) = P(f(x) \geq f(x^+) + \theta) = \phi(\frac{\mu(x) - f(x^+) - \theta}{\sigma(x)}) \tag{28}$$

$\Phi(\cdot)$ is normal distribution function; $f(x^+)$ is optimal objective function value at present; $\mu(x)$, $\sigma(x)$ is a posterior distribution of $f(x)$. When θ=0, it means θ tends to astringe to the $f(x^+)$, and can avert falling into the local optimal situation availably.

4.2. Relevance Vector Machine

Relevance Vector Machine (RVM) is a learnable data features model. In this model, training set are composed of inputs $\{x_n\}_{n=1}{}^N$ and corresponding target values $\{t_n\}_{n=1}{}^N$. The mapping equation in RVM is:

$$y(x,\omega) = \sum_{i=1}^{N} \omega_i K(x,x_i) + \omega_0 \tag{29}$$

$k(x,x_i)$ is kernel function; ω_i is corresponding weight; and N is the amount of training samples.

Assuming $\{t_n\}_{n=1}{}^N$ is autocephalous stochastic variable, the conditional probability of t is:

$$p(t\left|\omega, \sigma^2\right.) = (2\pi\ \sigma^2)^{-\frac{N}{2}} \exp(-\frac{\|t - \Phi\omega\|^2}{2\sigma^2}) \tag{30}$$

$t = (t_1 t_2 \ldots t_N)^T$, Φ is the kernel function matrix.

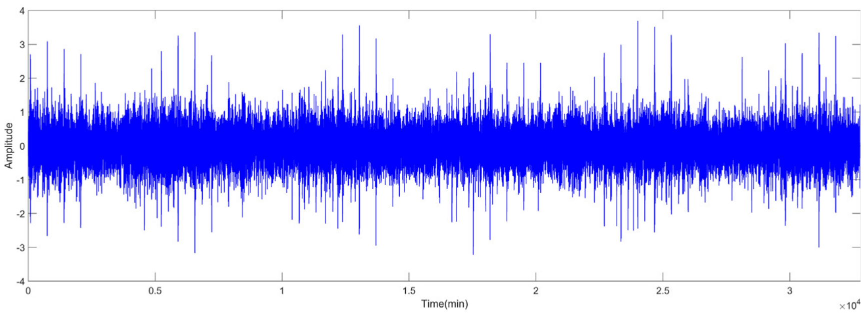


Fig. 8. Signal after ITD noise reduction.

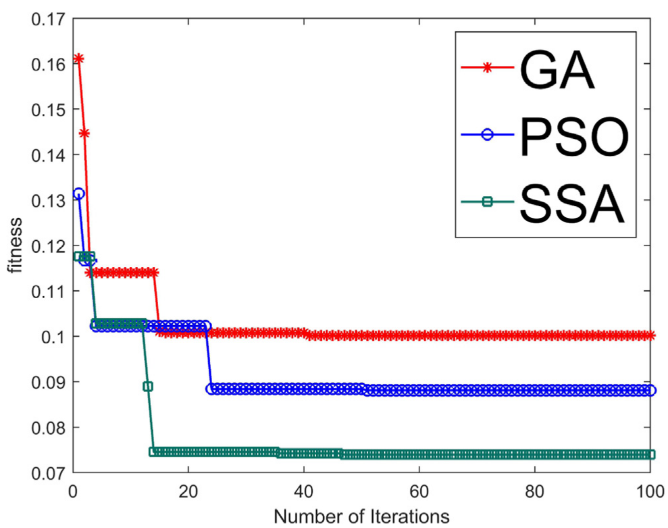


Fig. 9. Convergence trend of fitness values of MCKD parameters optimized by three methods.

To achieve sparsity, assumed $\omega_n \sim N(0, 1/\alpha_i)$, then conditional probability distribution of ω is:

$$p(\omega \Big| \alpha) = \prod_{i=1}^{N} \frac{\alpha_i}{\sqrt{2\pi}} exp(-\frac{\alpha_i {\omega_i}^2}{2}) \tag{31}$$

Using the maximum likelihood estimation to get the optimal solution of $p(t|\alpha, \sigma^2)$ is expressed as α_{MP} and ${\alpha_{MP}}^2$ respectively. After getting the optimal solution α_{MP} and ${\alpha_{MP}}^2$, the prediction can be made according to the weight parameters.

However, kernel functions and parameter can impact the performance of RVM model. Therefore, we can choose two or more kinds of functions, and uses weighted integrate them. The linear array of it can integrate the advantages of each kernel function and effectively advance forecast accuracy of RVM model. In this study, Gaussian kernel function and Laplace kernel function are used as kernel function, two selected functions are combined according to weight coefficients. The detail of kernel function is:

$$\begin{cases} K(x,x_i)_1 = exp(-\frac{\|x-x_i\|^2}{2d^2}) \\ K(x,x_i)_2 = exp(-\frac{\|x-x_i\|}{d}) \\ K = pK_1 + (1-p)K_2 \end{cases} \tag{32}$$

p is the weight coefficient of kernel function, $0 \le p \le 1$, $p = 0$ or $p = 1$ are single kernel function respectively; d is the width of the kernel.

4.3. BOA-MKRVM prediction model

The core parameters that influence the performance of MKRVM model are arguments and weights in the kernel function. In many cases, these two parameters are not optimal and have low accuracy. Therefore, BOA was selected to optimize kernel function parameters and weights of MKRVM, get a faster and more accurate BOA-MKRVM model. And basic steps are shown in Fig. 3:

Table 1
Detailed bearing parameters.

Parameter	Value	Parameter	Value
Outer race diameter	39.80 mm	Inner race diameter	29.30 mm
Bearing mean diameter	34.55 mm	Ball diameter	7.92 mm
Number of balls	8	Contact angle	0°
Load rating (static)	6.65 kN	Load rating (dynamic)	12.82 kN

5. Experiment analysis

Dataset utilizzato

5.1. Experimental data

Dato

The life cycle data of bearings were furnished by Xi 'an Jiaotong University-Shenyang Joint Laboratory of Science and Technology (XJTU-SY). The test facility include AC motor, motor speed controller, support bearing, a test bearing and so on. A PCB352C33 transducer is put in the horizontal and vertical orientations of bearing to gather the vibration signal. The total of 32 768 data points are recorded in the first 1.28 s within 1 min. The test bench is shown in Fig. 4. The type for experiment is LDK UER204, and parameters are expounded in Table 1. Since the radial force is applied in the horizontal direction, choose the vibration signal in this direction for the experiment, which can better represent the process of bearing degradation.

5.2. Detailed procedures

The experimental process is shown in Fig. 5:

5.3. Noise removal based on ITD-AMCKD

To verify the availability of ITD-AMCKD come up with in practical working conditions, horizontal data of bearing 3–2 is selected as experimental data in Fig. 6.

Firstly, ITD decomposition was carried out for the noisy signal, and a total of 9 PR components were obtained. Next, the PR evaluation index K proposed in this paper is used to screen the decomposed PR. The concrete value contained in the evaluation index in Table 2, based on the mean of the evaluation index K of 1.29, so retained the PR component with K greater than the mean. After calculation, it is inferred that PR1, PR2 and PR3 components after decomposition contain most of the effective information of the signal in Fig. 7. The signal reconstructed by the PR1, PR2 and PR3 component is shown in Fig. 8. From the figure that the signal constituted by the components screened by the evaluation indexes can effectively reduce signal noise.

Nel documento della doppia rete neurale non era chiaro il meccanismo utilizzato per il denoising!

In questo caso sembra funzionare molto bene

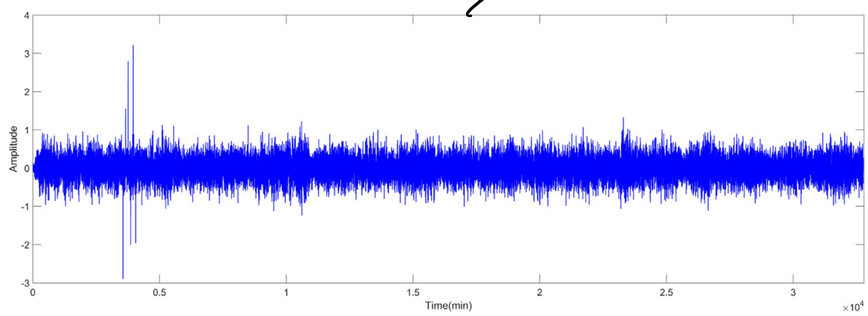


Fig. 10. Signal after secondary noise reduction.

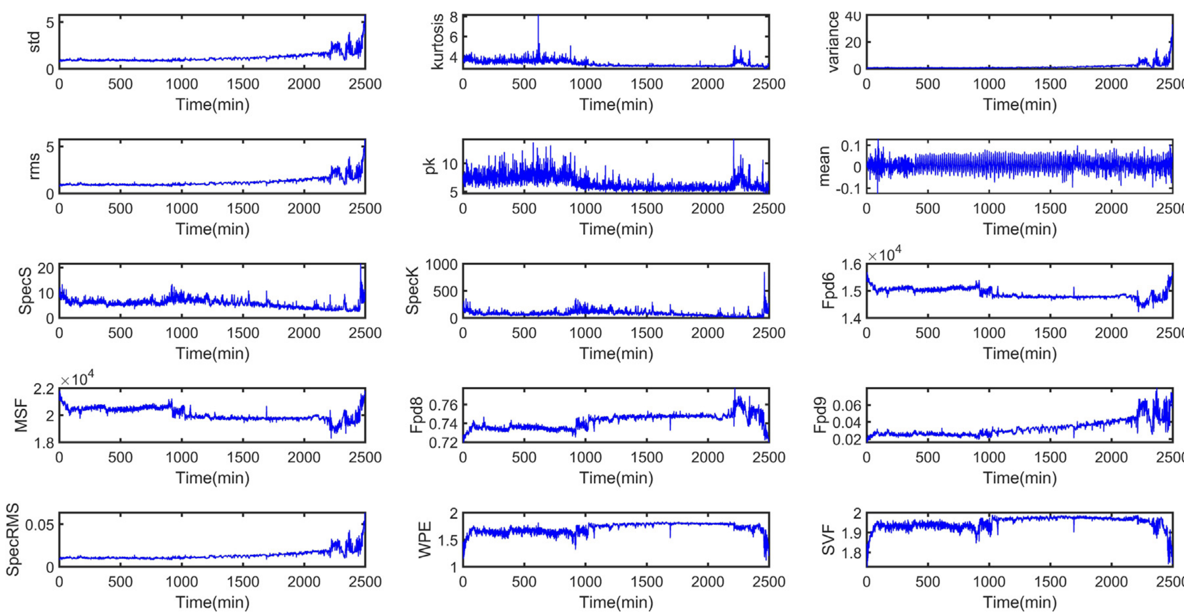
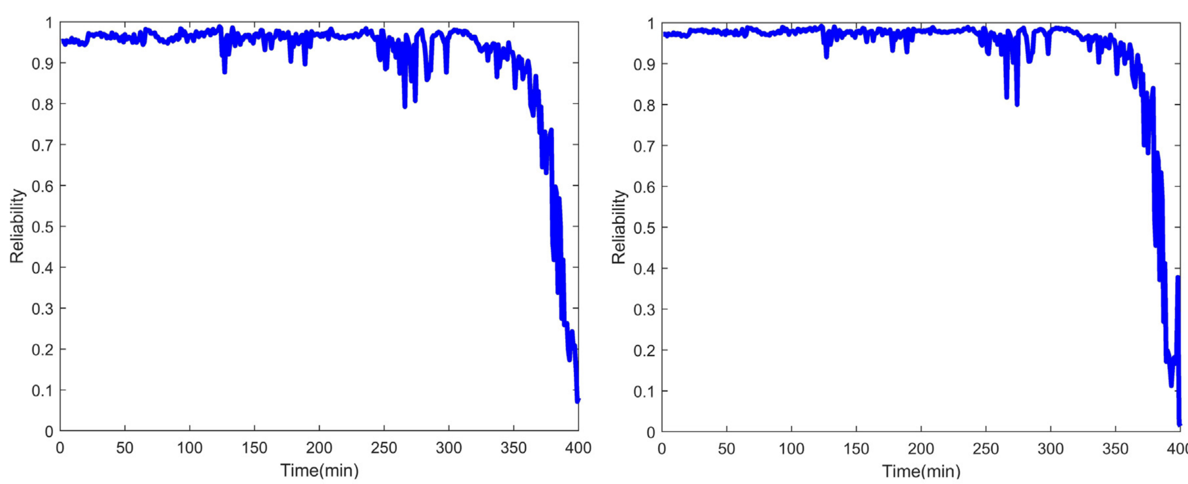


Fig. 11. Time-domain, frequency-domain and timely frequency-domain characteristics.



(a) LR

(b) WD

Sparrow Search Algorithm

Fig. 12. Real curve of bearing reliability.

Next, AMCKD algorithm was used to carry out secondary denoising of the effective signal, and parameters of MCKD were optimized by SSA algorithm. The range of M and L is set as [1,5] and [50,400] respectively. The initial sparrow number is 20, and iterative times is 30. The best displacement number M and filter length L are 1 and 398 respectively. To verify the accuracy of SSA proposed in this paper in MCKD parameter optimization, PSO and GA were used to optimize parameters of MCKD. Fig. 9 shows the convergence trend of SSA, PSO and GA. The optimal fitness value of SSA appears in the 14 iteration. In comparison to classical optimization method, SSA has fast optimizing velocity and high precision.

The signal after secondary denoising by ITD-AMCKD has less burr than the original signal, as demonstrated in Fig. 10. In general, the signal-to-noise ratio (SNR) is used to indicate the amount of the noise

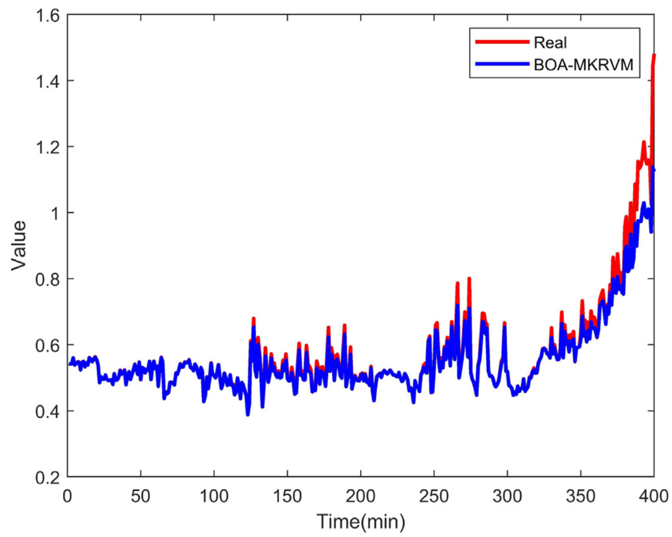


Fig. 13. BOA-MKRVM predicts the degradation state.

Table 2
PR evaluation index.

Index	RCMDE	P	RMSE	K
PR1	4.40	0.72	0.88	3.62
PR2	5.85	0.75	0.87	5.01
PR3	5.72	0.35	0.85	2.31
PR4	4.50	0.04	0.73	0.23
PR5	3.16	0.06	0.67	0.21
PR6	2.61	0.07	0.85	0.22
PR7	2.17	0.01	0.88	0.02
PR8	0.95	0.01	0.88	0.01
PR9	1.82	9.80e−04	0.88	0.01

Table 3
Comparison of Signal SNR.

Method	SNR
ITD	−3.1579
LCD	−3.1240
LMD	−3.0324
ITD-MCKD	−2.6822
ITD-AMCKD	−2.0449

signal in the signal, is the ratio of the noise in the signal. SNR value more higher show the signal have better quality, and on contrary, the worse. Find out from Table 3, the proposed method has the lowest SNR compared to other noise reduction algorithms, indicating that the use of ITD-AMCKD can decrease noise signals accurately.

5.4. Reliability prediction based on BOA-MKRVM

Fifteen time-domain, frequent-domain and timely frequent-domain characteristics, such as mean, kurtosis, standard deviation, waveform factor, spectral skewness, spectral kurtosis and mean square frequency were extracted from the denoised rolling bearing vibration signals. Timely frequent-domain characteristic wavelet packet entropy and singular value factor, as shown in Fig. 11. Multi-domain feature extraction consists of bearing feature parameter sets, but directly substituting the parameter sets into the logistic regression model will bring great difficulties to modeling. PCA often used to map high-dimensional data to low-dimensional data. The principle is to select the dimensions that can maximize the sample variance, that is, a small number of comprehensive indicators to represent the information existing in each vector. First, by composing the data into a matrix, and then each row of the matrix is zero-mean, also get the covariance matrix of it. The eigenvector of the covariance matrix is arranged into a matrix from top to bottom according to the size of the corresponding eigenvalues. The first r rows are taken to form the matrix, which is the matrix after dimension reduction. We selected the first three principal components as dimension reduction results with the cumulative contribution rate of 95% as the standard. The three principal components as the eigenvector.

Table 4
Error comparison of prediction models.

Model	3-2		3-4		5	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
BOA-MKRVM	0.027	0.028	0.020	0.015	0.019	0.016
BOA-RVM	0.032	0.056	0.029	0.046	0.028	0.020
MKRVM	0.096	0.112	0.087	0.124	0.075	0.050
RVM	0.258	0.390	0.170	0.242	0.168	0.246
LSSVM	0.082	0.083	0.075	0.065	0.047	0.068
BP	0.068	0.086	0.036	0.079	0.069	0.048
ELMAN	0.056	0.070	0.037	0.065	0.054	0.061
ELM	0.055	0.083	0.029	0.034	0.035	0.043

The eigenvector set was selected as the degradation characteristic information as argument of logistic regression model to get real curve of bearing reliability in Fig. 12(a). In addition, the eigenvector set is selected as the degenerate feature information, and the base function is used as the Weibull distributed proportional fault model to build the reliability curve in Fig. 12(b). After comparing the two figures at 390 points, it can be seen that the logistic regression model is more realistic than the reliability curve obtained by the Weibull proportional fault model. This is due to the many parameters of the Weibull proportional fault model, which is prone to problems in solving the parameters, and can easily lead to errors in reliability evaluation. Therefore, we establish the bearing reliability evaluation model by using the logistic regression model.

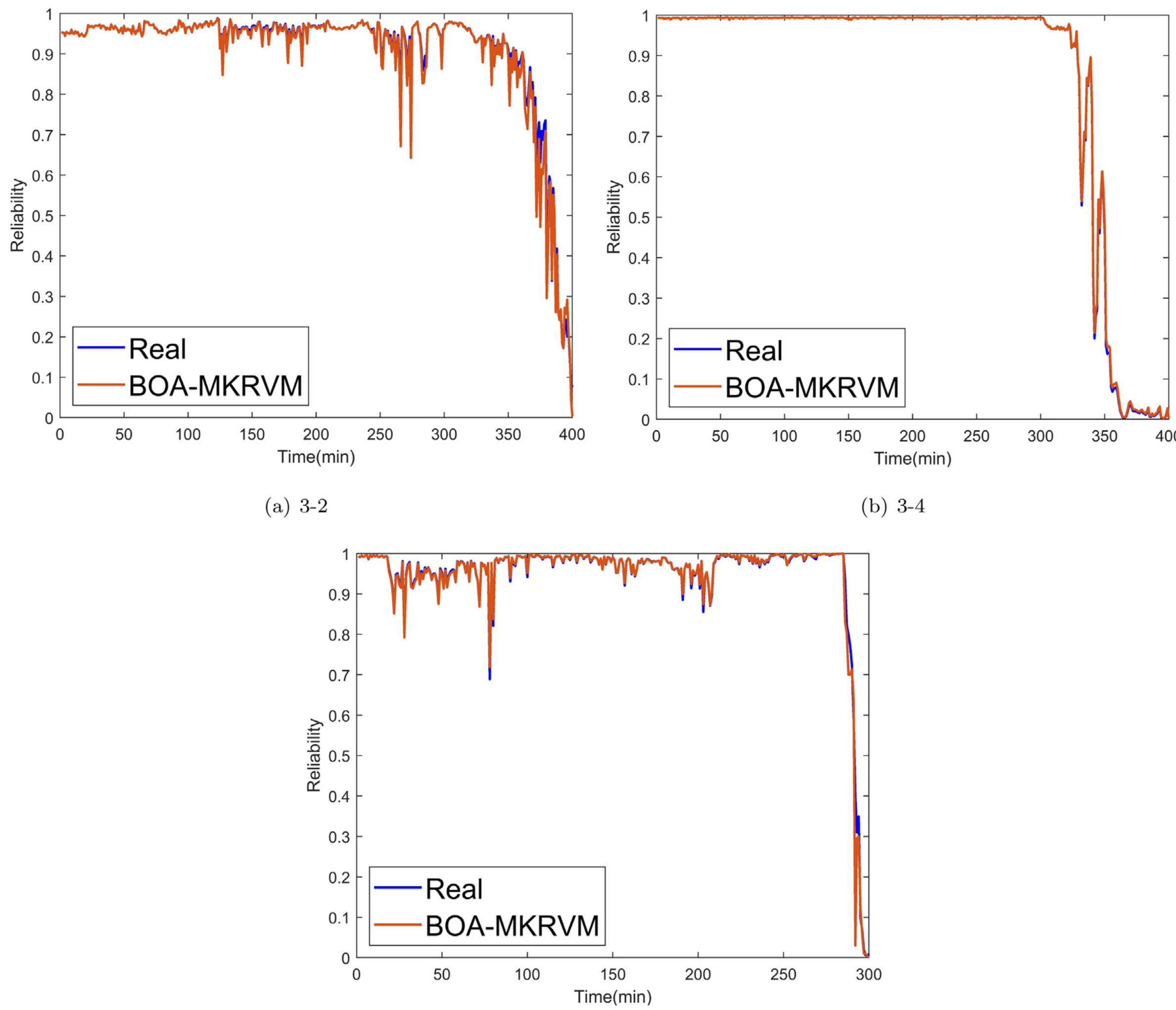
Next, the reliability prediction of bearings is carried out, in this study, firstly, utilize BOA to optimize kernel function parameters and weights of MKRVM. The optimal kernel function parameter and weight is 0.857, 1.327, take these results as parameters to MKRVM. Next, select 2096 training data and 400 test data in BOA-MKRVM to make predictions. Then, the final result of BOA-MKRVM are put into logistic regression model to get operation reliability of bearings. The comparison predicted degradation state and real data is demonstrated in Fig. 13.

To confirm accuracy of BOA-MKRVM is given in this study, the use of BOA-RVM, MKRVM, RVM, LSSVM, ELM, ELMAN, ELMBP, BP neural network forecasting model to forecast the bearing reliability and comparing with BOA-MKRVM method. Meanwhile, the proposed method was validated using bearing 3–4 and the University of Cincinnati bearing vibration data bearing 5. Reliability curve of the three bearings in Fig. 14. The errors generated by each prediction model are shown in Table 4.

By looking into the figure above, compared with existing bearing reliability prediction methods, the curve obtained by BOA-MKRVM is closer to practical reliability curve, showing good prediction performance. As can be seen from Table 4, MAE, MAPE of BOA-MKRVM model are 0.027, 0.028 in bearing 3–2, all of which are the smallest compared with the errors generated by existing prediction methods, indicating the errors of BOA-MKRVM model are smaller and prediction accuracy is further advanced.

6. Conclusion

In reliability assessment and prediction, feature vectors after denoising should save the effective message of raw data and display changes of signals. Firstly, noise signal is decomposed by ITD to get components PRs. On the one hand, an evaluation index is proposed to screening PRs. On the other hand, improved MCKD algorithm is combined to carry out secondary denoising of signals and reserve effective information



(a) 3-2

(b) 3-4

(c) 5

Fig. 14. Reliability prediction results.

of bearing signals. Next, BOA was used to optimize key parameters of mixed kernel RVM, and BOA-MKRVM model was used to as and predict the bearing reliability. In this study, the data of bearings from Xi 'an Jiaotong University were used to verify the results. The results indicate this model has very conspicuous consequence on noise removal of signals and can well characterize signal characteristics, improves the prediction accuracy effectively.

CRediT authorship contribution statement

Shuzhi Gao: Validation, Formal analysis, Investigation. **Yifan Yu:** Conceptualization, Methodology, Software, Writing – original draft, Visualization. **Yimin Zhang:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research was supported by the National Key R & D Program of China (2019YFB2004400), the Liaoning Province High-end Talent Construction Project-Distinguished Professor of Liaoning Province, China (No. [2018]3533), Key (General) project of Liaoning Provincial Department of Education, China (LJKZ0435).

References

Ali, J.B., Chebel-Morello, B., Saidi, L., Malinowski, S., Fnaiech, F., 2015. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mech. Syst. Signal Process. (56-57), 150–172.

Attoui, I., Oudjani, B., Boutasseta, N., Fergani, N., Bouraiou, A., 2020. Novel predictive features using a wrapper model for rolling bearing fault diagnosis based on vibration signal analysis. Int. J. Adv. Manuf. Technol. 106 (9–12), 3409–3435.

Azami, H., Rostaghi, M., Abásolo, D., Escudero, J., 2017. Refined composite multiscale dispersion entropy and its application to biomedical signals. IEEE Trans. Bio-Med. Eng. 99, 2872–2879.

Bao, Z., Zhang, G., Xiong, B., Gai, S., 2020. New image denoising algorithm using monogenic wavelet transform and improved deep convolutional neural network. Multimedia Tools Appl. 79 (1), 7401–7412.

Bi, F., Li, L., Zhang, J., Ma, T., 2015. Source identification of gasoline engine noise based on continuous wavelet transform and EEMD–RobustICA. Appl. Acoust. 100 (15), 34–42.

Bian, B.B., 2017. Fault diagnosis method of rolling bearing based on LCD denoising and LS-SVM. Modul. Mach. Tool Autom. Manuf. Tech. 2, 119–122.

Ding, F., Zhang, X., Wu, W., Wang, Y., 2020. A fault feature extraction method of motor bearing using improved LCD. IEEE Access 8, 220973–220979.

E, J., Ye, J., He, L., Jin, H., 2021. A denoising carbon price forecasting method based on the integration of kernel independent component analysis and least squares support vector regression. Neurocomputing 434 (2), 67–79.

Fan, H., Shao, S., Zhang, X., Wan, X., Cao, X., Ma, H., 2020. Intelligent fault diagnosis of rolling bearing using FCM clustering of EMD-PWVD vibration images. IEEE Access 8, 145194–145206.

Far, M.S., Mahsuli, M., Khoojine, A.S., Hosseini, V.R., 2021. Time-variant reliability-based prediction of COVID-19 spread using extended SEIVR model and Monte Carlo sampling. Results Phys. 26, 104364.

Frei, M.G., Osorio, I., 2007. Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals. Proc. Math. Phys. Eng. Sci. 463 (2078), 321–342.

Gao, S., Zhang, S., Zhang, Y., Gao, Y., 2020. Operational reliability evaluation and prediction of rolling bearing based on isometric mapping and NoCuSa-LSSVM. Reliab. Eng. Syst. Saf. 201, 106968.

Han, M., Wu, Y., Wang, Y., Liu, W., 2021. Roller bearing fault diagnosis based on LMD and multi-scale symbolic dynamic information entropy. J. Mech. Sci. Technol. 35 (5), 1993–2005.

Hao, R., Xuejun, A.N., Shi, Y., 2019. Single-channel blind source separation based on EMD and CICA and its application to gearbox multi-fault diagnosis. J. Vib. Shock 38 (8), 225–233.

He, Z., Chen, G., Hao, T., Teng, C., Cheng, Z., 2020. Weak fault detection method of rolling bearing based on testing signal far away from fault source. J. Mech. Sci. Technol. 34 (3), 1035–1048.

Hou, J., Wu, Y., Gong, H., Ahmad, A.S., Liu, L., 2020. A novel intelligent method for bearing fault diagnosis based on EEMD permutation entropy and GG clustering. Appl. Sci. 10 (1), 386–402.

Huang, C.G., Yin, X., Huang, H.Z., Li, Y.F., 2019. An enhanced deep learning-based fusion prognostic method for RUL prediction. IEEE Trans. Reliab. (99), 1–13.

Islam, S., Yavari, E., Rahman, A., Lubecke, V.M., Boirc-Lubecke, O., 2018. Separation of respiratory signatures for multiple subjects using independent component analysis with the JADE algorithm. In: 40th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'18).

Li, W., Chen, J., Li, J., Xia, K., 2021. Derivative and enhanced discrete analytic wavelet algorithm for rolling bearing fault diagnosis. Microprocess. Microsyst. 82, 103872.

Li, H., Huang, G., Ou, J., Zhang, Y., 2020a. Intelligent fault diagnosis of planetary wheel bearing using constrained independent component analysis and stacked sparse autoencoder. In: ASME Turbo Expo 2020: Turbomachinery Technical Conference and Exposition.

Li, Q., Zuo, M.J., Liang, S.Y., 2020b. False Lipschitz penalty sparse low-rank matrix and chaotic bionic optimization for prognosis of bearing degradation. IEEE Trans. Reliab. 99, 1–17.

Liu, X., Zhang, X., Luan, Z., Xu, X., 2019. Rolling bearing fault diagnosis based on EEMD sample entropy and PNN. J. Eng. 23, 8696–8700.

Malhotra, A., Minhas, A.S., Singh, S., Zuo, M.J., Kankar, P.K., 2021. Bearing fault diagnosis based on flexible analytical wavelet transform and fuzzy entropy approach. Mater. Today: Proc. 43 (7), 629–635.

Mao, W., He, J., Zuo, M.J., 2019. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. IEEE Trans. Instrum. Meas. 69 (4), 1594–1608.

Mcdonald, G.L., Zhao, Q., Zuo, M.J., 2012. Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection. Mech. Syst. Signal Process. 33, 237–255.

Meng, Z., Li, J., Yin, N., Pan, Z., 2020. Remaining useful life prediction of rolling bearing using fractal theory. Measurement 156, 107572.

Miao, F., Zhao, R., 2020a. A new fault diagnosis method for rotating machinery based on SCA-FastICA. Math. Probl. Eng. 2020 (2), 1–12.

Miao, F., Zhao, R., Wang, X., Jia, L., 2020. A new fault feature extraction method for rotating machinery based on multiple sensors. Sensors (Basel, Switz.) 20 (6), 1713.

Ning, S., Han, Z., Wang, Z., Wu, X., 2016. Application of sample entropy based LMD-TFPF de-noising algorithm for the gear transmission system. Entropy 18, 414–426.

Pan, Z., Meng, Z., Chen, Z., Gao, W., Shi, Y., 2020. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. Mech. Syst. Signal Process. 144, 106899.

Qiu, G., Y, G., Chen, J., 2019. Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model. Measurement 150, 107097.

Rostaghi, M., Azami, H., 2016. Dispersion entropy: A measure for time-series analysis. IEEE Signal Process. Lett. 23 (5), 610–614.

Sun, W., Cao, X., 2020. Curvature enhanced bearing fault diagnosis method using 2D vibration signal. J. Mech. Sci. Technol. 34 (6), 2257–2266.

Wang, F., Liu, X., Deng, G., 2019. Remaining life prediction method for rolling bearing based on the long short-term memory network. Neural Process. Lett. 50, 2437–2454.

Wang, Y., Peng, Y., Zi, Y., Jin, X., Tsui, K.-L., 2016. A two-stage data-driven-based prognostic approach for bearing degradation problem. IEEE Trans. Ind. Inf. 12 (3), 924–932.

Wang, F., Wang, B., Dun, B., Li, H., Hang, Q., Zhun, H., 2018. Rolling bearing reliability evaluation based on improved logistic regression model. J. Vib., Measure. Diagn. 38 (1), 123–129.

Wang, Y., Yang, M., Li, Y., Xu, Z., Fang, X., 2021. A multi-input and multi-task convolutional neural network for fault diagnosis based on bearing vibration signal. IEEE Sens. J. 21 (9), 10946–10956.

Xu, Y., Cai, Z., Cai, X., Ding, K., 2019. An enhanced multipoint optimal minimum entropy deconvolution approach for bearing fault detection of spur gearbox. J. Mech. Sci. Technol. 33 (3–4), 2573–2586.

Xue, J., Shen, B., 2020. A novel swarm intelligence optimization approach: sparrow search algorithm. Syst. Sci. Control Eng. 8 (1), 22–34.

Yang, J., Chang, Y., Gao, T., Wang, J., 2020. Failure prediction of the rotating machinery based on CEEMDAN-apen feature and AR-UKf model. Appl. Sci. 10 (6), 2056–2074.

Shuzhi Gao

Yifan Yu

Yimiu Zhang