Rolling Bearing Fault Trend Prediction Based on Composite Weighted KELM

He Yu

Mechanical Engineering College, Shijiazhuang 050003, PR China.

Hong-ru Li[*]

Mechanical Engineering College, Shijiazhuang 050003, PR China.

Zai-ke Tian

Mechanical Engineering College, Shijiazhuang 050003, PR China.

Yu-Kui Wang

Department of Aviation Four-Station, Air Force Logistics College, Xuzhou 221000, PR China.

The abilities of different degradation feature types to characterize rolling bearing fault trend are distinctive. And even the characteristic ability of the same degradation feature can change at various times. Thus these feature samples possess heteroscedasticity. However, traditional kernel extreme learning machine (KELM) model assumes that different input samples' effects on the predicted value are equal, which results in low prediction precision and low computing efficiency. To solve this problem, a novel composite weighted KELM (CWKELM) prediction model, which is fused with explicit weighting and implicit weighting, is proposed. In both feature type scale and sample time scale, the feature samples and the prediction model are weighted according to the prediction error. An adaptive mutation particle swarm optimization (AMPSO) algorithm is applied in optimizing the penalty factor and the kernel parameter in the model. Taking various entropy features as the input samples, the proposed model is adopted to conduct one-step and multi-step prediction for rolling bearing fault trend. Experimental results show that this prediction model has higher prediction accuracy and computing efficiency compared with the traditional KELM model.

1. INTRODUCTION

Rolling bearing is one of the key components that are widely used and affect the health status of rotating machinery.[1] In order to prevent bearings and equipment from failure or damage, it's of great safety significance and economic value to carry out rolling bearing fault trend prediction. The key of rolling bearing fault trend prediction is to extract accurate degradation characteristics and establish a good predictive model.[2]

On the one hand, as the input sample of the prediction model, the degradation characteristics need to be sensitive and robust to the degradation state of the rolling bearings in the whole life cycle.[3] The nonlinear complexity characteristics based on the information entropy theory can measure the probability distribution difference of variable bearing vibration signals in different degradation stages from multiple perspectives. And these degradation characteristics can reveal the developing trend of the ball bearing degradation state in essence. Therefore, entropy characteristics, including multiscale entropy,[4] energy spectrum entropy,[5] singular spectrum entropy and spatial information entropy[6] etc. have been widely utilized in rotating machinery fault diagnosis and prediction. These characteristics are combined with time-frequency analysis methods and perform well in fault diagnosis and degradation state identification. Zhang et al[7] proposed a novel hybrid bearing fault classification method based on permutation entropy (PE) and ensemble empirical mode decomposition (EEMD) which can calculate the multi-scale intrinsic characteristics as the fault classification features. In order to extract accurate fault features from vibration signals, Li et al[8] employed multi-scale permutation entropy (MPE) to characterize the complexity of the product function (PF) components which are computed by local mean decomposition (LMD). Zheng et al[9] put forward a new and effective bearing fault diagnosis methodology based on fuzzy entropy (FuzzyEn) and a self-adaptive time-frequency analysis method named local characteristic-scale decomposition (LCD) which deals with rolling bearing vibration signals. In summary, the information entropy features are mostly applied in rolling bearing fault classification and diagnosis, and the effect is significant. However, there are few applications of entropy features in the characterization and prediction of the whole degradation life of rolling bearings. This is mainly due to the degradation state of rolling bearings changing at every moment in their full life cycles. Bearing vibration signals have a large amount of non-stationary and unbalanced data that are difficult to process. When dealing with degradation data during the bearings' whole life cycles, there are wide performance variations of different single entropy features on both type scale and time scale and it can very likely result in a large prediction error. Therefore, it's better to carry out an effective weighted fusion pro-

---
[*]Corresponding author.

cessing of different entropy features.

On the other hand, the current fault prediction theories are mainly divided into three categories including physical prediction model, data-driven prediction methodology and fusion prediction methodology.[10] Among them, physical prediction models utilize mathematical models such as damage rules to describe failure mechanism and failure modes. However, these damage rules are either linear or multi-linear and that implies physical models fail to meet the mechanical failure prediction requirements. Moreover, physical prediction models rely heavily on the analysis of physical mathematical model which is difficult to establish for rotating machinery with complex structure and multiple fault modes. Therefore, the physics-based prediction approach makes the prognosis a risky task. With the rapid development of computer science and artificial intelligence (AI), the data-driven prediction methodologies have covered a great number of new technologies and AI algorithms such as time series prediction model, particle filtering, regression analysis, hidden Markov model (HMM), artificial neural network (ANN), support vector machine (SVM) and extreme learning machine (ELM) etc.[11] The main idea of the date-driven prediction methodology is to employ data from past operations and current bearing conditions in order to predict the bearing fault trend or even forecast the remaining useful life (RUL).[12] It is interesting to find out that there are both connections and differences among ANN, SVM and ELM. And these three kinds of methodologies have been widely studied and utilized in the area of mechanical prognostics.[13–15] ELM model is a learning machine based on single hidden layer feedforward neural networks which is proposed by Huang G.B. When using this model for predicting, the only thing you should do is to set the node number of hidden layer neurons, and the input layer weight and the implicit layer deviations are randomly generated. Furthermore, the weight of the output layer can be obtained, and the whole process is completed through one training without iteration. Therefore, this model has a faster computing speed compared with ANN model. However, due to the random generation of some parameters of the model, the generalization ability is insufficient. Meanwhile, the prediction performance is susceptible to parameters such as the number of implicit nodes, and the stability is relatively poor as a result. Huang et al in 2012 borrowed the kernel mapping from SVM model to replace the random mapping in ELM model and put forward the concept of kernel extreme learning machine (KELM) to improve the deficiency of the original ELM model.[16] And the proposed new model has been widely studied and utilized in time series prediction in the past five years, and its prediction ability is stronger than SVM.[17, 18] However, there are few application cases of KELM model in terms of rolling bearing condition monitoring or fault prognostics. The main reason is that the rolling bearing monitoring data has strong nonlinear and non-stationary characteristics and is imbalanced in both date type scale and time changing scale. Han and Wang[19] indicate that the KELM model ignores that different prediction window sample points have a different influence upon the prediction performance and these sample points exhibit heteroscedasticity which can be harmful to the fault prognostics. Furthermore, Han and Wang[19] point out that it's necessary to assign different weights to the sample points in time scale to improve the prediction accuracy. This significant standpoint shows us a new road to the higher prediction accuracy.

Based on the above analysis, this paper proposes a novel bearing fault trend prediction model named composite weighted KELM (CWKELM). On the basis of the KELM structure, the training input samples come from three kinds of entropy features and these features are explicit weighted in the feature fusion layer. And the model structure is implicit weighted in the model optimization layer at the same time. In the proposed model, the prediction error of CWKELM is reduced by balancing the data structure in both feature type scale and time scale. Further more, explicit weighting can reduce the input samples' dimension and the calculation efficiency of the prediction model can be improved. An adaptive mutation particle swarm optimization (AMPSO) algorithm is employed to optimize the key parameters such as penalty factor and kernel parameter in the model in order to improve the prediction accuracy further. The optimized CEKELM model is then applied in bearing fault trend prediction based on bearing vibration monitoring signals. Several kinds of information entropy features extracted from vibration signals are taken as the input samples and the root mean square (RMS) which monitors the bearing's whole life from normal state to failure is taken as the predicted target value. After these steps, single step and multi-step prediction of bearing fault trend is conducted at last. The advantages of this model are verified in both predicting accuracy and computational efficiency compared with traditional models such as ELM, KELM, SVM and BP neural networks.

The rest of this paper is organized as follows. The basic principles of CWKELM model are explained in Section 2. In detail, Section 2.1. presents the process of the entropy feature extraction and the explicit weighting. Later, the KELM methodology and the implicit weighting are exhibited in Section 2.2. Furthermore, Section 2.3. gives the specific steps of CWKELM for rolling bearing fault trend prediction. Section 3 describes the bearing fatigue life test and explains the experimental data used in this paper. Section 4 shows how the model parameters are optimized with AMPSO algorithm. The predicted results with the novel model and other methodologies are discussed and analysed in Section 5. And the conclusion is reached in Section 6.

2. PRINCIPLES OF CWKELM

2.1. Entropy Feature Extraction and Explicit Weighted Fusion

The first step of forecasting the rolling bearing degradation state is to extract different kinds of information entropy characteristics as the input samples of the model from vibration signals during bearings' whole life. Of these entropy characteristics, multi-scale entropy (MSE), energy entropy, and singular entropy have a better performance and have been applied in many cases. The calculation process of MSE is detailed by Li et al.[20] The fault developing trend of rolling bearings can be better characterized by describing how far the rolling bearing monitoring state deviates from the normal state with relative entropy theory. Therefore, relative energy entropy (REE) and relative singular entropy (RSE) are defined on the basis of energy spectrum entropy and singular spectrum entropy as follows:

One single set of vibration data at rolling bearing normal

state $x_b(t)(t = 0, 1, 2, \ldots, N-1)$ and M sets of vibration data at rolling bearing degradation state $x_i(t)(i = 1, 2, \ldots, M, t = 0, 1, 2, \ldots, N-1)$ are selected and then time-frequency spectrum $L_b(t, f)$ and $L_i(t, f)$ are computed by time-frequency analysis. And these two spectrums' energies at characteristic frequency f_g are respectively E_b^g and E_i^g calculated as follows:

$$E_b^g = \sum_{t=0}^{N-1} L_b(t, f_g)^2, \quad (1)$$

$$E_i^g = \sum_{t=0}^{N-1} L_i(t, f_g)^2, \quad (2)$$

where $g = 1, 2, \ldots, G$ represents the frequency characteristics of fault. Based on this, the energy spectrum of samples at normal state and samples at fault state can be represented as $E_b = [E_b^1, E_b^2, \ldots, E_b^G]$ and $E_i = [E_i^1, E_i^2, \ldots, E_i^G]$.

The total energy at characteristic frequency of all the samples can be computed by the following formula:

$$E_L^g = \sum_{i=1}^{M} E_i^g + E_b^g. \quad (3)$$

The proportions of the energy of $x_b(t)$ and $x_i(t)$ at characteristic frequency f_g accounting for the total energy are respectively as below:

$$p_b^g = E_b^g / E_L^g, \quad (4)$$

$$p_i^g = E_i^g / E_L^g, \quad (5)$$

where $p_b^g + \sum_{i=1}^{M} p_i^g = 1$.

According to the theory of relative entropy, REE between the fault sample set and the normal sample set is calculated as follows:

$$REE_b^i = \sum_{g=1}^{G} |p_i^g \log \frac{p_i^g}{p_b^g}|. \quad (6)$$

The time-frequency spectrum $L_b(t, f)$ of normal state and $L_i(t, f)$ of fault state are processed by singular value decomposition (SVD). The singular value spectrum of normal samples is $\sigma_b = [\sigma_b^1, \sigma_b^2, \ldots \sigma_b^J,]$ and the single value spectrum of fault samples is $\sigma_i = [\sigma_i^1, \sigma_i^2, \ldots \sigma_i^J,]$ where J is the order of the diagonal matrix in the decomposition of singular values.

Combined with singular spectrum entropy theory, two corresponding probabilities are defined as below:

$$q_b^j = \sigma_b^j / \sigma_L^i, \quad (7)$$

$$q_i^j = \sigma_i^j / \sigma_L^i, \quad (8)$$

where $1 \leq j \leq J, \sigma_L^j = \sum_{i=1}^{M} \sigma_i^j + \sigma_b^j$.

And RSE between the fault sample set and the normal sample set is calculated as follows:

$$RSE_b^i = \sum_{j=1}^{J} |q_i^j \log \frac{q_i^j}{q_b^j}|. \quad (9)$$

The above information entropy features can form a three-dimensional eigenvector: $X_1 = [\text{MSE REE RSE}]$. Three kinds of information entropy features measure the complexity changing of vibration signals from different angles and there

is much difference in abilities of characterizing the bearing fault trend. It will result in large prediction error that X_1 is taken as the input samples of KELM directly. Therefore, in the stage of model training, each entropy feature's prediction error for bearing operating states is evaluated respectively and conduct explicit weighting on every entropy feature. The weighted input samples are more suitable for prediction model training and fault prognostics. Relative Root Mean Square Error (RRMSE) and Correlation Coefficient of Trend Change α are two different kinds of indicators commonly used in evaluating the forecasting results. The actual feature sequence is set as $Y = \{y_j \mid j = 1, 2, \ldots, n\}$ and the predicted feature sequence is written as $Y_1 = \{\widehat{y}_j \mid j = 1, 2, \ldots, n\}$. The formulas of RRMSE and α are as follows:

$$RRMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(\frac{y_j - \widehat{y}_j}{y_j} \right)^2} \times 100\%, \quad (10)$$

$$\alpha = \frac{\sum_{j=1}^{n-1} [(y_{j+1} - y_j)(\widehat{y}_{j+1} - \widehat{y}_j)]}{\sqrt{\sum_{j=1}^{n-1} [(y_{j+1} - y_j)(\widehat{y}_{j+1} - \widehat{y}_j)]^2}}. \quad (11)$$

RRMSE indicates the deviation and α reflects the consistency between the predicted values and the actual values respectively. Based on this, the weight evaluation factor is defined as follows:

$$F = \alpha / RRMSE. \quad (12)$$

The merits of single entropy feature prediction are evaluated by calculating the weight evaluation factors of single feature prediction results. And finally the input predicting features of the prediction model is expressed as the following formula.

$$X_2 = [\text{MSE REE RSE}] [F_{mse} F_{ree} F_{rse}]^T. \quad (13)$$

2.2. KELM and Implicit Weighting

The ELM model is a kind of single hidden layer feedforward neural networks and its output function can be represented as

$$f(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta. \quad (14)$$

In order to guarantee the minimum prediction error of output values, the corresponding objective function equation is set as follows:

$$min \parallel f(x) - f_0(x) \parallel$$
$$= min \parallel \sum_{i=1}^{L} \beta_i h_i(x) - f_0(x) \parallel = 0 \quad (15)$$

In the above equations, L is the number of neurons in the hidden layer. $h(x)$ is the random mapping in the hidden layer and $f_0(x)$ is the predictive function value. When there are N inputs named $x_i(i = 1, 2, L, \ldots, N)$, the matrix form of Eq. 14 can be represented as follows:

$$f(x) = H\beta, \quad (16)$$

where H is the random mapping matrix in the hidden layer. And matrix H can be denoted by $h(x)$ as follows:

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ M \\ h(x_N) \end{bmatrix}. \quad (17)$$

The output weight vector β can be calculated by the least square method as the following equation.

$$\beta = H^+ O = H^T (HH^T)^{-1} O$$
$$= H^T \left(\frac{I}{C} + HH^T \right)^{-1} O, \qquad (18)$$

where H^+ represents the Moore-Penrose generalized inverse matrix of H and O represents the predictive expectation matrix. C is the penalty factor that is generally positive and this parameter can improve the generalization ability of the model. And the output function of the model can be indicated as follows:

$$f(x) = H\beta = HH^T \left(\frac{I}{C} + HH^T \right)^{-1} O. \qquad (19)$$

However, unlike ELM, KELM model uses the kernel mapping in the hidden layer instead of random mapping in ELM, which can be shown as the following formula 17.

$$\Omega_{ELM} = HH^T : \Omega_{ELM,i,j}$$
$$= h(x_i) \cdot h(x_j) = K(x_i, x_j), \qquad (20)$$

where $K(x_i, x_j)$ is a kind of kernel function, which is always selected as RBF kernel function. The general form is given as follows:

$$K(x_i, x_j) = \exp \left(-\frac{\| x_i - x_j \|^2}{2\sigma^2} \right), \qquad (21)$$

where σ is the kernel parameter. And the output function of KELM model can be represented by the following equation.

$$f(x) = \begin{bmatrix} K(x, x_1) \\ M \\ K(x, x_N) \end{bmatrix}^T (I/C + \Omega_{ELM})^{-1} O. \qquad (22)$$

On the time scale, the predictive ability of the models input samples is different. That is to say, the closer the prediction points are, the greater the impact of training samples on prediction results will be.[19] Therefore, there is necessity conducting implicit weighting on the input samples of KELM model on the time scale. Concretely, weighting coefficients are used to weigh the prediction error variables of the model, and then the objective function 15 can be rewritten as follows:

$$min P(\beta, \xi) = \frac{1}{2} \beta^T \beta + \frac{1}{2} C \sum_{i=1}^{N} \nu_i \xi_i^2, \qquad (23)$$

$$s.t. h(x_i)\beta = o_i - \xi_i, i = 1, L, \ldots, N,$$

where $P(\beta, \xi)$ is the cost function of the prediction model. x_i is the value of the input samples. o_i is the prediction expectation value. ξ_i is the error variance. N is the number of training samples. Combined with Eq. 17 and Eq. 18, the output function of the implicit weighted KELM model is shown as follows.

$$f(x) = \begin{bmatrix} K(x, x_1) \\ M \\ K(x, x_N) \end{bmatrix}^T (V_C + \Omega_{ELM})^{-1} O, \qquad (24)$$

where $V_C = diag \left\{ \frac{1}{C\nu_1}, L, \frac{1}{C\nu_N} \right\}$ and the weighting coefficient is $\nu_k = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^{N} \xi_k^2}}{|\xi_k|}$.

2.3. The Steps of CWKELM Prediction Methodology

Based on the above principles of the explicit weighting in entropy feature extraction and the implicit weighting in KELM prediction, the CWKELM prediction steps are determined as follows:

(1) Signal preprocessing and feature extraction

The information entropy features and the condition monitoring data are extracted and normalized from the rolling bearing life test.

(2) Parameter optimization

AMPSO algorithm is used to optimize the kernel parameter and the penalty factor of the prediction model.

(3) Explicit weighting

The weight evaluation factors of three different kinds of entropy features are calculated and explicit weighting is conducted according to Eq. 13.

(4) Implicit weighting

The implicit weighting is conducted on the input samples according to the weighting coefficient in the proposed model.

(5) Fault trend prediction

The proposed model, CWEKLM, is utilized to predict the rolling bearing operation monitoring data. And the specific step flow chart is shown in Fig. 1.

3. FATIGUE LIFE TEST FOR ROLLING BEARINGS

In this paper, the fatigue test data of rolling bearings that are analysed by the methodology mentioned above are gathered in Bearing Test Research Center, Hangzhou, China.[21] As shown in Fig. 2, the experimental platform is mainly composed of three parts, including ABLT-1A bearing test machine, signal acquisition module and status monitoring module. Figure 3 is the system diagram. Four CA-YD-139 acceleration sensors are arranged in the four bearing test stations, and connected with the DH-5920 dynamic signal acquisition instrument. In this way, four sets of rolling bearings can be tested at one time and multiple sets of full life vibration data can be stored in the end. Meanwhile, four thermal resistances on stations and one YD-1 acceleration sensor are connected with one signal amplifier to monitor the operation indexes such as temperature, kurtosis and RMS. When these indexes exceed the alarm threshold, the test machine will stop working.

The single row deep groove ball bearing 6204, which is commonly used in mechanical equipment, is taken as the test object of the life test, and it is shown in Fig. 4(a). During the test, the motor speed is 1500 r/min, and the vibration signals are sampled every 10 minutes. The sampling time is 1 s, and the sampling frequency is 25.6 kHz. When the test runs to 9600 minutes, the kurtosis index is over eight while it is about three at normal state. The rolling bearing has a serious failure and the test machine stops. After the shutdown and examination, the bearing of No.4 Station fails due to the inner ring erosion as is shown in Fig. 4(b) and 960 groups of rolling bearing vibration data is collected.

The change curve of Root Mean Square (RMS) over time is shown in Fig. 5. This monitoring index monitors and records the whole process of 6204 bearing from normal state to failure
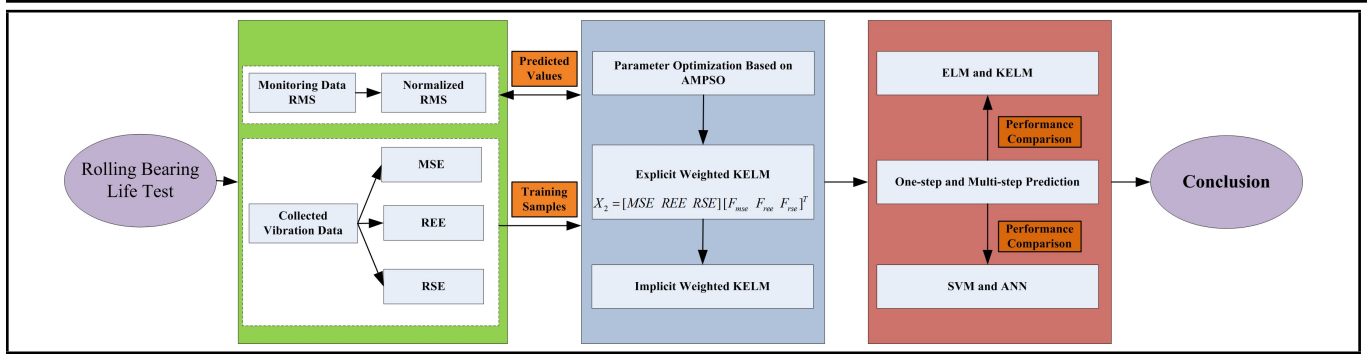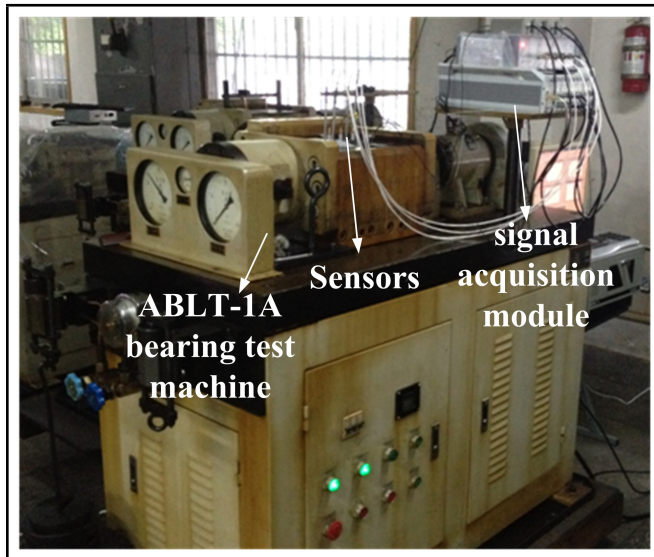
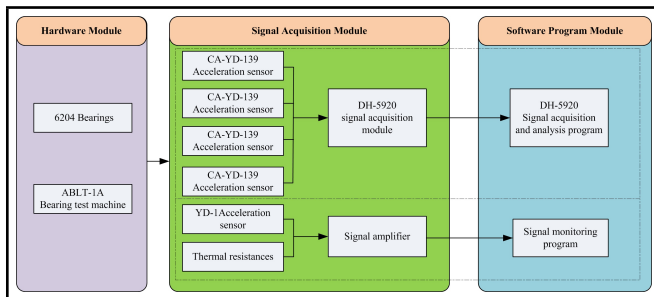**Figure 1.** Flow chart of the methodology.

Figure 2. Bearing test platform.

Figure 3. Structure of the test system.

Figure 4. a) 6204 Bearings at normal state. b) 6204 Bearing with inner ring pitting.

Figure 5. The RMS-time curve.

state and has a good ability to follow and reflect the fault developing trend of the bearing. Therefore, predicting the fault trend of the rolling bearing can be equivalent to predicting the RMS monitoring curve. According to the change of curve of time, the fault development of rolling bearing over time can be divided into four stages: (1) normal state:0–6820 min; (2) slight degradation state:6830–8450 min; (3)serious fault state:8460–9330 min; (4) failure state:9340–9600 min. Liu et al[22] point out that the second and the third state which are referred as "soft failure" account for most of mechanical failures. Considering that there is some regularity in time sequence of the period, we take 6840-9330 min as the analysis interval of the prediction model where the training interval is 6840–8830 min, 200 sets of data in total; the test range is 8840–9330 min, 50 sets of data in total.
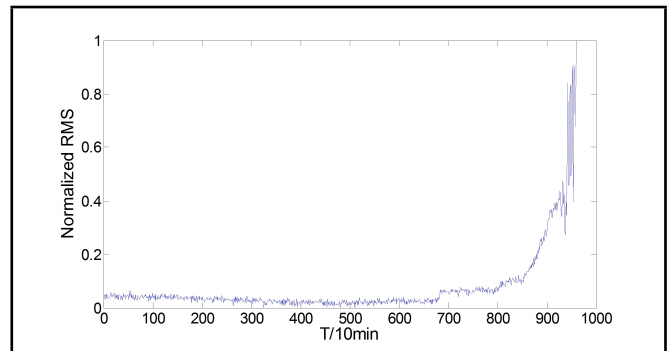
4. OPTIMIZATION OF THE MODEL PARAMETERS

Similar to the support vector machine (SVM) model, the penalty factor C in KELM model is mainly used to balance the complexity of the model and the empirical risk value to improve the generalization performance of the model. In order to control the model complexity, the factor C is always relatively small but cannot be too small, avoiding experience error.[22] When the kernel parameter is too large, the response velocity of the model can be slow and the ability to adjust is poor. While the kernel parameter σ is too small, the model can be too sensitive and this can cause a great error too. Therefore, it's necessary to optimize the penalty factor C and the kernel parameter σ. At present, commonly used optimization algorithms include Grid Search Algorithm, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm.

Compared with the Grid Search Algorithm and GA, PSO algorithm converges fast and has strong universality but its search precision is poor and it is easy to get into local optimal. However, GA has relatively high precision despite relatively complex calculation and low iteration efficiency. Therefore,
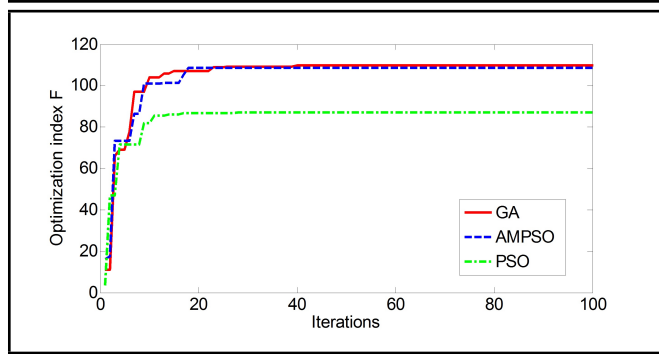
Figure 6. The iteration curve of parameter optimization.

the adaptive mutation particle swarm optimization (AMPSO) algorithm is proposed to be applied to C and σ optimization. The main idea of this algorithm is to add the variation operation to the particle variable and initialize the population to the optimal solution.[23] The training period data is used for the single step prediction analysis. The number of particle group dimension is set to 2, and the number of particles is set to 20, and the number of iterations is limited to 100. $C \in [0.1, 1000]$, $\sigma \in [0.01, 100]$. The fitness objective function is the weight evaluation factor F. Parameter setting and specific steps of the algorithm can be seen in Qu et al[23] and Zhang.[24] At the same time, GA and PSO algorithm are used for comparison and further analysis. The detail specification of the calculating device is given as below: CPU is Intel®Core™ i5-4590 CPU@3.30GHz, RAM is 4.00GB, and the operating system (OS) is 32-bit Windows 7. With this calculating device, the converge curves are shown in Fig. 6. The optimization results are shown in Table 1.

The AMPSO algorithm achieves the optimal value of 108.5, which is basically consistent with the optimal solution of GA, only through eighteen iterations. The computational efficiency of AMPSO is higher than GA. The PSO algorithm, which doesn't add the variation operation, is in the local optimum after the 11th iteration, and the optimization result is not ideal. Finally the model parameters are selected as $C = 9.8$, $\sigma = 1.02$.

5. THE ANALYSIS AND COMPARISON OF FAULT TREND PREDICTION

There are 200 sets of data in the training period that are used to train the prediction model. The information entropy features are weighted by the one-step prediction weighting evaluation factor. Firstly, the information entropy characteristics including MSE, REE and RSE are extracted from the vibration signals, and then three groups of feature samples are respectively used as the model input to predict the changing curve of RMS over time. The prediction results are shown in Figure 7, 8, 9. The predicted values which are calculated from MSE, REE and RSE deviate from the real values more or less in the second half of the training period. There are relatively more error points in the predicted curve which is calculated from MSE. However, the error points are fewer in the predicted curve which is calculated from RSE. The three information entropy features' abilities to predict the RMS curve differ from each other and the training results need further quantitative analysis.

The RRMSE and Correlation Coefficient of Trend Change α of the above mentioned one-step prediction training results are
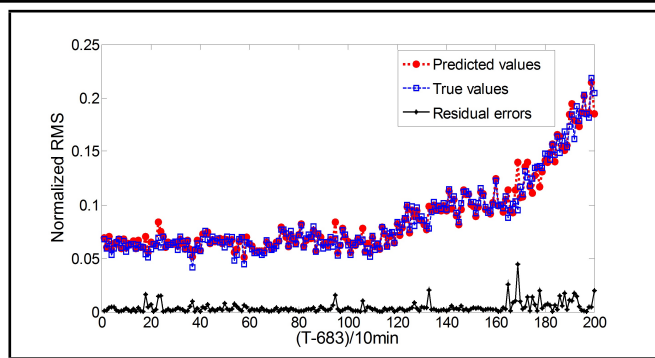

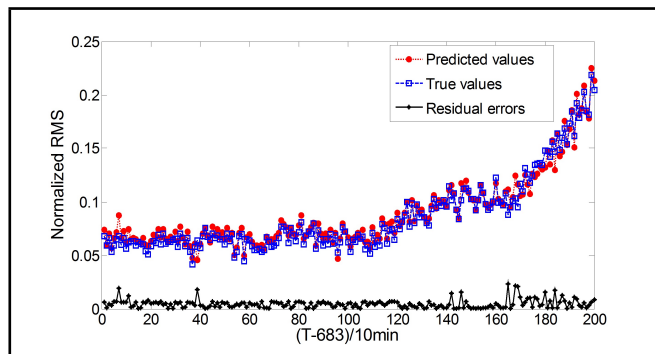**Figure 7.** The MSE training result of one-step forecast.

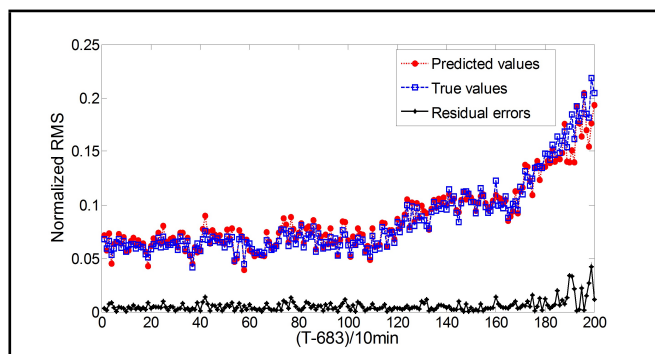**Figure 8.** The REE training result of one-step forecast.

Figure 9. The RSE training result of on-step forecast.

respectively calculated and the weight evaluation factor F is computed according to this. The results are shown in Table 2.

Among the three features, the correlation coefficient of RSE is the largest, and the RRMSE of RSE is similar to that of MSE. The weight evaluation factor of RSE is the largest, and that indicates the predictive ability of RSE is the best. This result is consistent with the prior qualitative analysis. Meanwhile, the indexes of RSE and REE are close to each other and all of them are within reason. That means it's reasonable and effective to utilize information entropy features for rolling bearing fault prognostics. On the other hand, in order to utilize the comprehensive entropy information in the samples, explicit weighting is conducted on the input samples of KELM model according to Eq. 13. The fusion feature X_2 is used as a new input characteristic to train KELM model. The training result is shown in Fig. 10.

After the explicit weighting processing, the predicted residuals in the middle and posterior segment of the single information feature are effectively cut down and the predicted error is further reduced. In order to illustrate the effectiveness of the

Table 1. Comparison of the optimization results.

Optimization Algorithm	Parameter C	Parameter σ	Target Value F	Iteration Number	Computation Time/s
AMPSO	9.8	1.02	108.5	18	12.235
GA	10.5	1.12	109	23	15.363
PSO	5.7	0.84	85.5	11	9.275

Table 2. The evaluation indexes of single entropy feature.

Information Entropy Features	Correlation Coefficient	RRMSE /%	Weight Evaluation Factor (F)
MSE	6.4430	7.54	85.4152
REE	6.7806	7.77	87.3206
RSE	7.4317	7.58	97.9829

Table 3. The comparison of fusion processings.

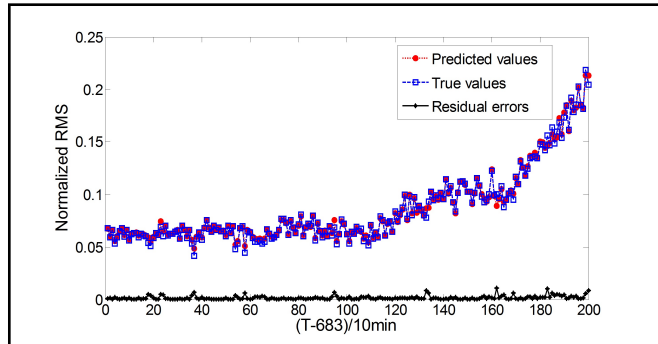| Fusion Processing	Correlation Coefficient	RRMSE /%	Weight Evaluation Factor (F)
Multi inputs (without processing)	7.0093	6.40	109.5788
Explicit weighting	7.6351	3.15	242.6733
Composite weighting	8.3976	1.68	501.3071

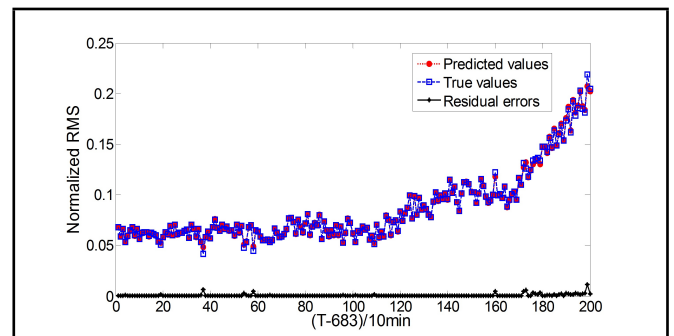**Figure 10.** The explicit weighted training result of one-step forecast.

Figure 12. The composite weighted training result of one-step forecast.
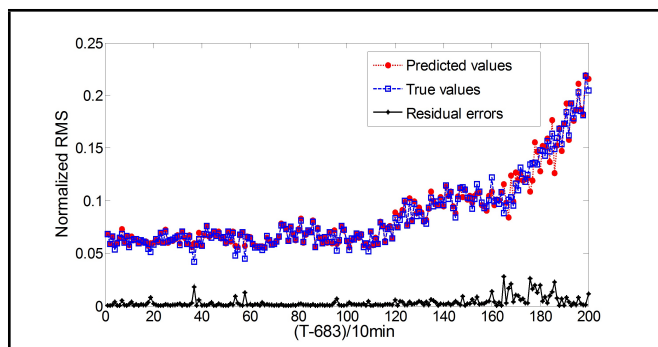

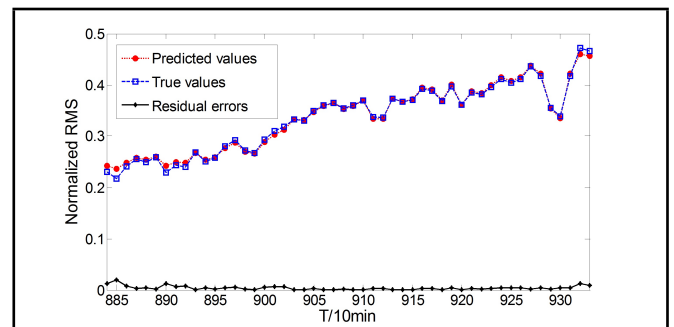**Figure 11.** The multi-input training result of one-step forecast.

Figure 13. CWKELM test result of one-step forecast.

explicit weighting, the three single entropy features which are not weighted are used as the KELM input samples for training and the training time is 0.2027s. However, the weighting fusion features training time is only 0.0913s. The one-step forecast result of the multi input features which are not weighted is shown is Fig. 11. Compared with this result, the predicted error of the explicit weighted training result in Fig. 10 is relatively smaller, and that demonstrates the effectiveness of the explicit training.

On the basis of explicit training, implicit weighting is conducted on the characteristic sequence X_2 on the time scale in KELM model and the result is shown in Fig. 12 below. Within the training interval, the predicted values are consistent with the true values and the error is relatively small. The further comparative analysis can be seen in Table 3.

By the comparative analysis of Table 2, Table 3 and Figure 12, we can conclude that the prediction accuracy is significantly improved after the composite weighting in KELM and training samples. The proposed model's prediction result has a high consistency with real value in the range of one-step forecast and the prediction curve fits the monitoring curve

well in the training stage. The CWKELM prediction model is used to predict 50 groups of data in the test period, as is shown in Fig.13. The KELM multi-input prediction model is adopted for comparison and the result is shown in Fig. 14. By comparing Fig. 13 and Fig. 14, we find that the prediction effect of CWKELM for rolling bearing fault trend is better than KELM. In the middle and posterior section of the test interval, the predicted curve is basically consistent with the true curve and there is only a small margin of deviation in the initial phase of the test interval. In comparison, the prediction error of KELM model is relatively large and the KELM model is sensitive to the change of the data samples. And this results that small amplitude of data fluctuation can cause a large prediction error. Therefore, the CWKELM prediction model has some engineering significance in balancing the data structure and reducing the prediction error.

In order to further illustrate the robustness and practical value of CWKELM model in rolling bearing fault trend prediction, a certain length of data is randomly selected as the training sample and ELM, KELM and CWKELM model are adopted to the one-step, five-step, ten-step, fifteen-step and
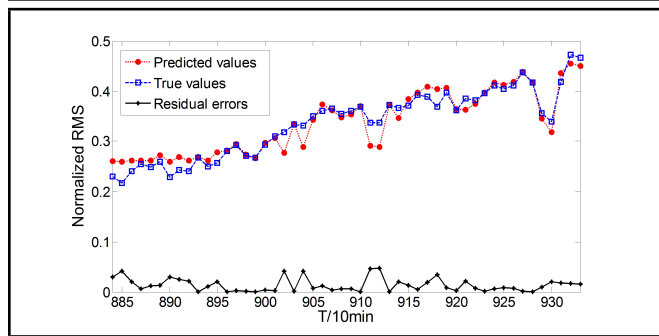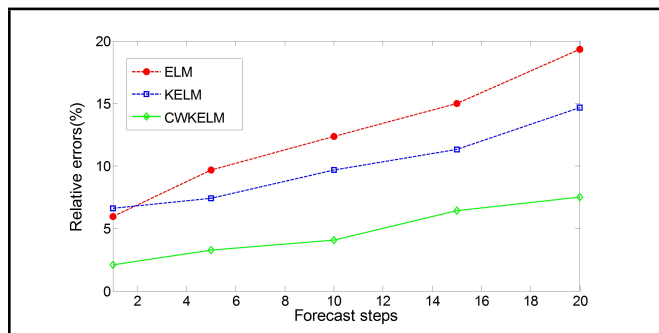
Figure 14. KELM test result of one-step forecast.

Figure 15. The multi-step forecast error curve.

Table 4. Performance comparison of prediction models.

Prediction Model	RRMSE /%	Correlation Coefficient	Computation Time/s
CWKELM	2.32	7.8546	0.4568
SVM	3.76	6.4533	0.5786
BP	5.27	5.9271	1.3624

twenty-step prediction of the normalized RMS. The result is shown in Fig. 15.

From Fig. 15, the relative errors of the three models have increased in different degrees with the improvement of the number of prediction step. Compared with the ELM and KELM models, the relative error of CWEKLM model is relatively small, and it is always below 10% and the performance of CWKELM model is better. While the relative error of ELM model and KELM model is over 15% when the number of prediction step reaches 20. The reliability of the predicted results is significantly reduced.

As a special kind of single hidden layer feedforward neural network, KELM combines the design concept of ANN and the kernel learning methodology of SVM, and has been widely applied in the field of fault diagnosis and prognostics. CWKELM is a kind of advanced KELM model. In order to illustrate the advantages of CWKELM in rolling bearing fault trend prediction, SVM regression prediction model and the typical back propagation (BP) neural network are utilized for comparison and the results are shown in Table 4.

Table 4 shows that the RRMSE of CWKELM model is relatively smaller than that of SVM and BP model and the fault trend changing correlation coefficient between the predicted results and the real values is larger. This result can demonstrate that the CWKELM model can better achieve the goal of rolling bearing fault trend prediction. In terms of computational efficiency, CWKELM prediction model has a shorter calculation time. The main reason is that the composite weighting changes

multi inputs into single fusion inputs and the structure is simple, so the computation speed is faster. However, there are always multiclass sample inputs in the SVM model and BP neural network and the computation process is complex and needs more time. In the control experiment, BP neural network takes empirical risk minimization as its principal and need a large number of training samples to train. However, due to the limited number of samples in the study, there will be a large error. CWKELM model can balance the sample data structure and modify the input samples in both sample data scale and time scale by explicit weighting and implicit weighting. That ensures the CWKELM model also has certain adaptability for small sample data.

6. CONCLUSIONS

In order to improve the rolling bearing fault trend prediction accuracy and computational efficiency, this paper proposes a composite weighted KELM (CWKELM) model and applies it to forecast the condition monitoring index RMS. Through comparison and analysis, the following conclusions can be reached.

(1) CWKELM model can overcome the differences of the input samples in both type scale and time scale and balance the data structure by explicit weighting and implicit weighting. In this way, the input samples' ability to characterize different degradation states can be enhanced. In the one-step prediction and multi-step prediction of rolling bearing fault trend, the CWKELM model is more accurate than traditional KELM model.

(2) Composite weighting can improve the prediction accuracy of the KELM model. At the same time, explicit weighting can convert multi inputs into single inputs and thus improve the computational efficiency to some degree. Compared with KELM, SVM and BP neural network, CWKELM model has faster computing speed.

(3) AMPSO algorithm is applied in the parameter optimization of CWKELM model and this algorithm can effectively avoid the local optimal problem which exists in PSO algorithm. Meanwhile, compared with GA, AMPSO algorithm has advantages such as fewer iterations, faster convergence speed and is more suitable for real-time online parameter optimization, and then the prediction accuracy of CWKELM model is improved rapidly.

7. AKNOWLEDGEMENTS

REFERENCES

[1] Wang, X. L., Gu, H., Xu, L., Hu, C., and Guo, H. A SVR-Based Remaining Life Prediction for Rolling Element Bearings, *J. Fail. Anal. Prev*, **15**(4), 548–554, (2015). https://dx.doi.org/10.1007/s11668-015-9976-x

[2] Shen, Z. J., Chen, X. F., He, Z. J., Sun, C., Zhang, X. L. and Liu, Z. W. Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine, *Chin. J. Mech. Eng*, **49**(2), 183–189, (2013). https://dx.doi.org/10.3901/JME.2013.02.183 (In Chinese)

3 Zhang, S., Zhang, Y. X., Li, L., and Zhu, J. P. Rolling Elements Bearings Degradation Indicator Based on Continuous Hidden Markov Model, *J. Fail. Anal. Prev*, **15**(5), 691–696, (2015). https://dx.doi.org/10.1007/s11668-015-9999-3

4 Liu, H. H. and Han, M. H. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings, *Mech. Mach. Theory*, **75**(5), 67–78, (2014). https://dx.doi.org/10.1016/j.mechmachtheory.2014.01.011

5 Ai, Y. T. and Fei, C. W. Rotor vibration fault diagnosis method based on wavelet energy spectrum entropy and SVM, *Journal of Aerospace Power*, **26**(8),1830–1835, (2011). https://dx.doi.org/10.13224/j.cnki.jasp.2011.08.026 (In Chinese)

6 Wang, Y. K., Li, H. R., Wang, B., and Xu, B. H. Spatial Information Entropy and Its Application in the Degradation State Identification of Hydraulic Pump, *Mathematical Problems in Engineering*, **2015**(7), 1–11, (2015). https://dx.doi.org/10.1155/2015/532684

7 Zhang, X. Y., Liang, Y. T., Zhou, J. Z., and Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM, *Measurement*, **69**, 164–179, (2015). https://dx.doi.org/10.1016/j.measurement.2015.03.017

8 Li, Y. B., Xu, M. Q., Wei, Y., and Huang, W. H. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, *Measurement*, **77**, 80–94, (2016). https://dx.doi.org/10.1016/j.measurement.2015.08.034

9 Zheng, J. D., Cheng, J. S., and Yang, Y. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy, *Mech. Mach. Theory*, **70**, 441–453, (2013). https://dx.doi.org/10.1016/j.mechmachtheory.2013.08.014

10 Jay, L., Wu, F. J., Zhao, W. Y., Masoud, G., Liao, L. X., and David, S. Prognostics and health management design for rotary machinery systemsReviews, methodology and applications, *Mechanical Systems and Signal Processing*, **42**(1), 314–334, (2014). https://dx.doi.org/10.1016/j.ymssp.2013.06.004

11 Farzaneh, A. and Jan, L. Remaining useful life estimation: review, *Int. J. Syst. Assur. Eng. Manag.*, **5**(4), 461-474, (2014). https://dx.doi.org/10.1007/s13198-013-0195-0

12 Idriss, E. and Erkki, J. A summary of fault modelling and predictive health monitoring of rolling element bearings, *Mechanical Systems and Signal Processing*, **60–61**, 252–272, (2015). https://dx.doi.org/10.1016/j.ymssp.2015.02.008

13 Lei, X., Chen, X. H., Zhang, X. H., and Liu, M. A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition. *Journal of Intelligent Manufacturing*, **28**(8),1893–1914, (2017). https://dx.doi.org/10.1007/s10845-015-1077-x

14 Huang, H. Z., Wang, H. K., Li, Y. F., Zhang, L. L., and Liu, Z. L. Support vector machine based estimation of remaining useful life: current research status and future trends, *Journal of Mechanical Science and Technology*, **29**(1),151-163, (2015). https://dx.doi.org/10.1007/s12206-014-1222-z

15 Liu, Y. B., He, B., Liu, F., and Zhao, J. W. Remaining Useful Life Prediction of Rolling Bearings Using PSR, JADE, and Extreme Learning Machine, *Mathematical Problems in Engineering*, **2016**, 1–13, (2016). https://dx.doi.org/10.1155/2016/8623530

16 Huang, G. B., Zhou, H. M., Ding, X. J., and Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification, *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, **42**(2), 513–529, (2012). https://dx.doi.org/10.1109/TSMCB.2011.2168604

17 Liu, X. W., Wang, L., Huang, G. B., Zhang, J., and Yin, J. P. Multiple kernel extreme learning machine, *Neurocomputing*, **149**, 253–264, (2015). https://dx.doi.org/10.1016/j.neucom.2013.09.072

18 Shahaboddin, S., Kasra, M., Chen, H. L., Ganthan, N. S., Dalibor, P., Ma, C. Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran, *Journal of Atmospheric and Solar-Terrestrial Physics*, **134**, 109–117, (2015). https://dx.doi.org/10.1016/j.jastp.2015.09.014

19 Han, M. and Wang, X. Y. Multivariate chaotic time series prediction based on weighted extreme learning machine, *Control Theory & Applications*, **30**(11), 1467–1472, (2013). https://dx.doi.org/10.7641/CTA.2013.30170 (In Chinese)

20 Li, L., Zhu, Y. L., and Song, Y. Q. Feature Research of Vibration Signal of Power Transformer Using Multiscale Entropy, *Journal of Vibration, Measurement & Diagnosis*, **35**(4), 757–762, (2015). https://dx.doi.org/10.16450/j.cnki.issn.1004-6801.2015.04.026 (In Chinese)

21 Li, H. R., Wang, Y. K., Wang, B., Xu, B. H., and Li, X. L. The method of grey Markov remaining service life prediction specific to generalized mathematical morphological, *Journal of Vibration, Measurement & Diagnosis*, **28**(2), 316–323, (2015). https://dx.doi.org/10.16385/j.cnki.issn.1004-4523.2015.02.019 (In Chinese)

22 Liu, S. J., Hu, Y. W., Li, C., Lu, H. T., and Zhang, H.C. Machinery condition prediction based on wavelet and support vector machine, *Journal of Intelligent Manufacturing*, **28**(4), 1045–1055, (2017). https://dx.doi.org/10.1007/s10845-015-1045-5

23 Qu, J., Chen, H. Y., Liu, W. Z., Zhang, B., and Li, Z. B. Application of Support Vector Machine Based on Adaptive Mutation Particle Swarm Optimization in Analysis of Gas Mixture, *Chinese Journal of Sensors and Actuators*, **28**(8),1262–1268, (2015). https://dx.doi.org/10.3969/j.issn.1004-1699.2015.08.027 (In Chinese)

24 Zhang, Y. W. and Zhang, P. C. Optimization of nonlinear process based on sequential extreme learning machine, *Chemical Engineering Science*, **66**(20), 4702–4710, (2011). https://dx.doi.org/10.1016/j.ces.2011.06.030