Fault diagnosis of wind bearing based on multi-scale wavelet kernel extreme learning machine

Siwen Zhu and Bin Jiao

Department of Electrical Engineering, Shanghai Dian Ji University, Electrical College, No. 300 blooms Road, Pudong New Area, Shanghai, China.
Email: 278630427@qq.com

Abstract. The principle of kernel Extreme Learning Machine (ELM) is demonstrated. On this basis, a multi-scale wavelet kernel extreme learning machine is proposed. The multi-scale wavelet kernel is used as the kernel function of the extreme learning machine. The test shows that it is an achievable extreme learning machine. Experiments show that, using the multi-scale wavelet kernel extreme learning machine in the wind turbine bearing fault diagnosis has higher classification accuracy and speed than the support vector machine classification algorithm, and has excellent application value.

1. Introduction
Wind turbines often work in harsh outdoor environments and have a high incidence of failures. As an important transmission parts of wind turbine, the bearing has the function of reducing the relative movement of the component and supporting the guide. Once the bearing broke out, it will cause the unit to stop running and bring great economic losses to the whole wind farm. Therefore, it is very important to quickly and effectively diagnose the failure of main bearing of wind turbine.

At present, many artificial intelligence algorithms are used in wind turbine bearing fault diagnosis. Vapnik and so on proposed support vector machine (SVM) on the basis of statistical theory [1], which effectively solved the problem of non-linear classification, but does not have a good classification for the limited sample and has long training time and low accuracy; On the basis of that he input layer connection weights of the feedforward neural network do not require iterative adjustment, Huang proposed a new neural network training architecture called Extreme Learning Machine (ELM) [2]. Compared with the traditional BP neural network based on gradient descent, this method can solve the least square solution of output weight at one time, and it is fast and generalized. It is widely used in classification problem. [3] Compared with the traditional classification method, ELM greatly reduces the training time on the basis of guaranteeing certain recognition accuracy [4]. ELM is faster and parameter-less, easier to deploy than SVM. [5]- [6] The method has good nonlinear mapping ability and is widely used in SVM. In Literature [7], the kernel method can also be applied to the limit learning machine algorithm. In Literature [9], the wavelet kernel function can be applied to the limit learning machine [8], and the performance is better. In this paper, the multiscale wavelet kernel function is applied to the ELM, and the fault diagnosis of the wind turbine bearing is carried out. The experiment shows that the method has high classification accuracy and speed.
2. ELM and its Improvements

2.1. ELM

N different training sample sets are \{ (x_i, t_i) | x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m \}. The activation function of the feedforward network is \( g(x) \) and the number of hidden layer nodes is L. ELM training steps are as follows:

1) Determine the feedforward neural network structure;
2) Randomly set the connection weight of the feedforward neural network \((a_i, \cdots, b_i)\) and calculate the output matrix \(H\) of the hidden layer;

\[
(a_i, \cdots, b_i), i = 1, \cdots, L
\]

Let \( h(x_i) = G(a_1 \cdot x_1 + b_1) \cdots G(a_L \cdot x_i + b_L) \)

\[
H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}_{N \times L}
\]

\( H \) is an \( N \times L \) matrix.
3) Solve the output weight of the least squares norm solution:

\[
\beta = H^T (\frac{1}{C} + HH^T)^{-1} Y
\]

Where I is the unit matrix and \( C \) is a constant. \( T \) is the expected matrix. The classification formula of the limit learning machine is shown below:

\[
f(x) = \text{sign}(h) \left( H^T \left( \frac{1}{C} + HH^T \right)^{-1} Y \right)
\]

2.2. Kernel Extreme Learning Machine (KELM)

Since the input weight of the ELM does not need to be adjusted during the training process, it can be given at once when the network is established. The implicit layer output of the ELM can be described as

\[
\sum_{i=1}^{L} \beta_i G(a_i \cdot x_1 + b_i) = t_1, i = 1, 2, \cdots, L
\]

\[
\sum_{i=1}^{L} \beta_i G(a_i \cdot x_N + b_i) = t_N
\]

Where \( G(x) \) is the activation function. \( a_i = [a_{i,1}, \cdots, a_{i,N}]^T \) is the input right. \( \beta_i \) is the output right. \( b_i \) is the offset of the \( i \)-th hidden layer unit. \( a_i \cdot x_N \) is inner product operation. Equation (5) can be reduced to a matrix form expression of equation (6)

\[
H \beta = T
\]

Where \( H \) is the implicit output layer and \( T \) is the desired output.

\[
H(a_1, \cdots, a_L, b_1 \cdots b_L, x_1 \cdots x_N) =
\begin{bmatrix} G(a_1 \cdot x_1 + b_1) \cdots G(a_L \cdot x_1 + b_L) \\ \vdots \\ G(a_1 \cdot x_N + b_1) \cdots G(a_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}
\]

SVM determines that the optimal classification surface is based on the principle of structural risk minimization, but in the process of solving the ELM, it is necessary to consider not only the minimization of the empirical error, but also the structural risk minimization. Thus, a compromise is made between the minimized output weight and the minimized error, and the construction formula is as follows:

\[
\min: \| \beta \|^2 \text{ and } \sum_{i=1}^{N} \| \beta \cdot h(x_i) - t_i \|
\]

Equation (8) is equivalent to equation (9):
min: \[ L = \frac{1}{2} ||\beta||^2 + \frac{1}{2} C \sum_{i=1}^{N} ||\xi_i||^2 \]  
\[ s.t.: h(x_i)\beta = t_i^T - \xi_i^T, i = 1,2,\cdots,N \]  
(9)

Where \( \xi_i = [\xi_{i,1}, \cdots, \xi_{i,L}] \) is the error between the network output value and the actual value which correspond to the training sample \( x_i \).

According to the KKT condition, the Lagrange function is used to solve the most problem. That is, the problem of Equation (9) can be equivalent to

\[
\min L_{DELM} = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^{N} ||\xi_i||^2 - \sum_{i=1}^{N} \sum_{j=1}^{M} (h(x_i)\beta_j) - t_i^T + \xi_i^T, i = 1,2,\cdots,N
\]

(10)

Where \( \beta_j \) is the weight of the connection hidden layer and the j-th output node. The corresponding optimization constraints of \( \beta = [\beta_1, \cdots, \beta_m] \) are

\[
\frac{\partial L_{DELM}}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum_{i=1}^{L} \alpha_i h(x_i) \rightarrow \beta = HH^T \alpha \quad (a)
\]

\[
\frac{\partial L_{DELM}}{\partial \xi_i} = 0 \rightarrow \alpha_i = C\xi_i, i = 1,2,\cdots,N \quad (b)
\]

\[
\frac{\partial L_{DELM}}{\partial \alpha_i} = 0 \rightarrow h(x_i)\beta_i - t_i^T + \xi_i^T = 0, i = 1,2,\cdots,N \quad (c)
\]

The hidden layer output matrix \( H \) is only related to the number of samples and the number of nodes of the hidden layer but nothing to do with the number of output nodes of sample.

11 (a) and 11 (b) are incorporated into Formula 11 (c)

\[
h(x_1)H^T C_1 - t_1^T + \xi_1^T = 0
\]

\[
\vdots
\]

\[
h(x_N)H^T C_N - t_N^T + \xi_N^T = 0
\]

(12)

All the formula outputs are merged together, let

\[
T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix}
\]

(13)

Then (13) can be combined into

\[
\left( I + \frac{1}{C} HH^T \right) \alpha = T
\]

(14)

In the end we can derive

\[
\beta = HH^T (HH^T + \frac{I}{C})^{-1} T
\]

(15)

ELM approximation function can be written as:

\[
f(x_i) = h(x_i)H^T (HH^T + \frac{I}{C})^{-1} T
\]

(16)

where \( H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}_{N \times L} \).

The hidden layer output \( h(x_i) \) of each sample can be regarded as a non-linear mapping of the sample \( x_i \), this mapping can be in the form of an added ax + b or RBF.
According to the theory of kernel function, implicit mapping can be constructed to replace the inner product of \( h(x) \) in equation (17), that is, we can construct kernel function instead of \( HH^T \), as shown below

\[
HH^T(i, j) = K(x_i, x_j)
\]

\[
HH^T = \Omega_{ELM} = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}
\]  

(18)

\[
h(x)H^T = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}
\]

Also,

\[
f(x) = h(x)H^T(HH^T + \frac{I}{C})^{-1}T
\]  

(19)

So the ELM solution formula can be written as

\[
f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left( \frac{I}{C} + \Omega_{ELM} \right)^{-1}T
\]  

(20)

Through the above steps, a kernel extreme learning machine is constructed. The kernel extreme learning machine has a more powerful function approximation than the traditional extreme learning machine, and the ability to deal with sub-linear classification is also stronger.

2.3. Multi-scale Wavelet Kernel Function (MKELM)

Similar to the support vector machine, if a kernel function can satisfy the Mercer condition, then the kernel function can be used as the kernel function of the kernel extreme learning machine. To simplify the algorithm, the SVM kernel is applied to the ELM.

Known as a mother wavelet function \( h(x) \), its scalability and translation factors are \( a \) and \( b \), then the wavelet basis function can be expressed as follows:

\[
h_{a, b}(x) = \sqrt{|a|}h(x^\frac{b}{a})
\]  

(21)

According to the theory of tensor product, a multidimensional wavelet function can be written as tensor product of multiple one-dimensional wavelet functions:

\[
h(x) = \prod_{i=1}^{n} h(x_i)
\]  

(22)

According to (22), a translation of the kernel function can be constructed:

\[
K(x, x') = K(x - x') = \sum_{i=1}^{n} \frac{x_i - x'_i}{a}
\]  

(23)

In this experiment, Morlet wavelet function is selected as \( h(x) = \cos(1.75x) \cdot \exp(-x^2/2) \), then its corresponding wavelet kernel function can be expressed as follows

\[
\text{wavelet kernel}(x, x') = \prod_{i=1}^{n} \left[ \cos(1.75 \times \frac{x_i - x'_i}{a}) \cdot \exp\left(-\frac{(x_i - x'_i)^2}{2a^2}\right) \right]
\]  

(24)
Wavelet kernel function not only has the characteristics of strong nonlinear mapping, but also inherits the characteristics of wavelet analysis on non-stationary input parameters. Therefore, the wavelet kernel extreme learning machine with wavelet kernel function can approximate arbitrary function with high precision, which is not available in traditional kernel function.

Although the kernel function has a strong mapping ability and non-linear classification ability, but in some complex cases, the kernel extreme learning machine constructed by one wavelet kernel function does not meet practical application requirements such as data heterogeneous or irregular, large sample size, sample uneven distribution and so on. So combining multiple kernel functions to get better results is an inevitable choice.

Usually a mixed kernel function can be constructed by the way of the superposition of different kernel functions

\[ K = a_1 \times K_1 + \cdots + a_n \times K_n \]  

(25)

For the wavelet kernel function itself has the ability to expand the scale, different scale a can construct a different wavelet kernel function.

The kernel ELM whose kernel function is wavelet kernel function is still the single-scale wavelet kernel ELM. In order to improve the function approximation ability with multi-scale characteristics, this paper proposes multi-scale wavelet kernel to improve the classification accuracy. Multi-scale wavelet kernel function is constructed according to the properties of multi-scale wavelet kernel as follows

\[
K = \prod_i^p \left[ \cos \left( 1.75 \times \left( \frac{x_i - x_i'}{a_1} \right)^2 \right) \exp \left( -\frac{(x_i - x_i')^2}{2a_1^2} \right) \right] + \\
\prod_i^p \left[ \cos \left( 1.75 \times \left( \frac{x_i - x_i'}{a_2} \right)^2 \right) \exp \left( -\frac{(x_i - x_i')^2}{2a_2^2} \right) \right] + \\
\cdots + \prod_i^n \left[ \cos \left( 1.75 \times \left( \frac{x_i - x_i'}{a_N} \right)^2 \right) \exp \left( -\frac{(x_i - x_i')^2}{2a_N^2} \right) \right]
\]  

(26)

Formula (26) can be abbreviated as

\[
K = \sum_i^p \prod_i^p \left[ \cos \left( 1.75 \times \frac{x_i - x_i'}{a_i} \right) \exp \left( -\frac{(x_i - x_i')^2}{2a_i^2} \right) \right]
\]  

(27)

Since the multiscale wavelet kernel function is a combination of wavelet kernel functions of multiple scales, the multi-scale wavelet kernel extreme learning machine (MKELM) can greatly relax the selection of the kernel parameters, and even can dilute the choice of kernel function, and its approximation ability higher, with the advantages of easily using, better recognition and so on.

3. Experimental Simulation

3.1. Simulation of Standard Data

In order to verify the validity of MKELM and the performance comparison with other kernel function ELM, this paper tests the standard classification test data of wine in UCI with Multi-scale wavelet kernel, wavelet kernel, gaussian kernel, poly kernel and no kernel. The test results are shown in Table 1; In order to verify the performance comparison between the MKELM and other algorithms, the standard data are classified by SVM, ELM, and MWKELM. Among them, SVM uses wavelet kernel function. The test results are shown in Table 2.

| Kernel functions           | Test accuracy | Test samples |
|----------------------------|---------------|--------------|
| Multi-scale wavelet kernel | 0.991         | 200          |
| Wavelet kernel             | 0.988         | 200          |
| Gaussian kernel            | 0.965         | 200          |
| Poly kernel                | 0.989         | 200          |
| No kernel                  | 0.967         | 200          |
As can be seen from Table 1, compared with other kernel functions, classification accuracy of the multi-scale wavelet kernel extreme learning machine is higher, the accuracy of no kernel and Gaussian is similar.

Table 2. Retrieval performance comparison of three classification algorithms

| Classification algorithms | Test accuracy | Test samples |
|--------------------------|--------------|--------------|
| SVM                      | 0.933        | 200          |
| ELM                      | 0.985        | 200          |
| MKELM                    | 0.991        | 200          |

As can be seen from Table 2, the ELM is more accurate than the SVM, and the MKELM has the best classification ability.

3.2. Simulation of Experiment

In order to further verify the effectiveness and superiority of MKELM in the fault diagnosis of wind turbine gearbox bearing, 25 sets of data were randomly selected as the training samples in 27 experimental data, and 21 sets of data were taken as test samples. The test results are shown in Table 3 Show.

Table 3. Comparison of experimental data test results

| Classification algorithms | Test accuracy | Test samples |
|--------------------------|--------------|--------------|
| SVM                      | 0.833        | 21           |
| ELM                      | 0.879        | 21           |
| MKELM                    | 0.905        | 21           |

As the Table 3 shows, the multi-scale wavelet kernel learning algorithm has the highest classification accuracy, the worst is the support vector machine SVM, the multi-scale kernel function method parameter selection complexity is much larger than the single kernel method parameter selection.[10] Therefore, the multi-scale wavelet kernel has better generalization ability. The accuracy of classification of multi-scale wavelet kernel learning machine is higher. However, the traditional ELM output weight is directly obtained by the least squares estimation method [11]. For the sake of learning robustness, by optimizing the selection parameters or solving the coherence of the training data and the gross error, such as the multi-hidden layer output matrix limit Learning machine algorithm [12], can further enhance the accuracy of learning machine

4. Conclusions

In this paper, a multi-scale kernel limit learning machine is proposed. The multi-scale wavelet kernel function is used as the kernel function of the extreme learning machine which satisfies the mercer condition and can be used as the allowable kernel of the extreme learning machine. In the application of the fault diagnosis of the gearbox bearing of the wind turbine, it is shown that the multi-scale wavelet kernel extreme learning machine has higher accuracy in fault classification, better performance than SVM and ELM, and has certain application value. However, the traditional ELM output weight is directly obtained by the least squares estimation method. For the sake of learning robustness, it is possible to further improve the accuracy of the learning machine by optimizing the selection parameters or solving the coherence of the training data and the gross error, such as the multi-hidden layer output matrix extreme learning machine. The selection of fault characteristics is also important and a more appropriate combination of features can be taken to get better retrieval results.

5. References

[1] Cortes C, Vapnik V. Support-Vector Networks [M]. Kluwer Academic Publishers, 1995.
[2] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications J. Neurocomputing, 2006, 70(1/3) 489-501.
[3] Wei W, Guang B Face recognition based on extreme learning machine J. Neurcomputing, 2011, 70(1/3) 489-501.
[4] Nahvi B, Habibi J, Mohammadi K, et al Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature J. Computers & Electronics in Agriculture, 2016,124 150-160.
[5] Zhang S, Song Y, Wang Z, et al Structural analysis for the joint of the ITER ELM coil J. Fusion Engineering & Design, 2017,114 47-51.
[6] Colton A L, Buttery R J, Fielding S J, et al ELM studies on the COMPASS-D tokamak J. Plasma Physics & Controlled Fusion, 2017,38(8) 1359-1365(7).
[7] Lu Y, Zhang W, Shi J, et al. Unsteady aerodynamic modeling of biaxial coupled oscillation based on improved ELM J. Aerospace Science & Technology, 2017,60 58-67.
[8] Fu H, Niu Z, Zhang C, et al. ASELM: Adaptive semi-supervised ELM with application in question subjectivity identification J. Neurocomputing, 2016, 207 599-609.
[9] Gunn J P, Carpentier-Chouchana S, Dejarnac R, et al .Ion orbit modelling of ELM heat loads on ITER divertor vertical targets J. Nuclear Materials & Energy.2016.
[10] Huang G B, Ding X, Zhou H. Optimization method based extreme learning machine for classification M. Elsevier Science Publishers B. V. 2010.
[11] Yadav B, Ch S, Mathur S, et al. assessing the suitability of extreme learning machines (ELM) for groundwater level prediction J. Journal of Water & Land Development, 2017, 32(1) 103-112.
[12] Kmiec’ K, Kot I Occurrence of aphids from Eriosomatinae subfamily on elms in the gree area in Lublin. J. Annales Umcs Horticultura, 2010, 20(1) 7-13.