Bearing fault detection in varying operational conditions using wavelet packet decomposition and support vector machine

Guozeng Liu¹, Jianmin Zhao and Haiping Li
Army Engineering University, Shijiazhuang, 050003, China

¹ E-mail: xiaoliu2005@sohu.com

Abstract. Roller bearing is one of the most important elements in machine components. In generally, the bearing can’t work in a steady condition. In this paper, we propose a method based on wavelet packet transform (WPT) and auto regressive (AR) model spectrum to extract fault features in varying operational conditions, and use support vector machine (SVM) to set an effective pattern recognition model. Our bearing experiment data comes from Case Western Reserve University Bearing Data Center. According to the result, WPT, AR model spectrum and SVM can solve the problem of fault diagnosis in varying operational conditions.

1. Introduction
Bearings play an important role in machine components in condition monitoring of rotating machineries. The faults of bearings always lead to machine breakdown. Therefore, the detection of faults in bearings is critical to engineer maintenance. In the process, feature extraction and pattern classification are vital to the detection. In recent decades, many methods of feature extraction and pattern classification have been developed in fixed work conditions [1, 2]. Extracting time domain and frequency domain feature from vibration signal is easy and convenient by Fourier transform. Researchers have summed up some time domain features like mean, standard deviation, root mean square, peak value, kurtosis, skewness etc. and frequency domain features like central frequency, power spectrum, mean frequency etc. [3, 4]. In fact, the analysis above is not suitable to deal with the non-stationary vibration signals. Then, researchers got some time-frequency domain analysis like short-time Fourier transform (STFT) [5], wavelets transform (WT) [6, 7] etc. Although the methods above are able to deal with non-stationary vibration signals, they still have difficulty detecting faults in bearings in varying operational conditions. In this case, researchers try to use wavelet packet transform (WPT) model to extract the main fault features and exclude the operational condition features [8-10]. Moreover, auto regressive (AR) model spectrum is a kind of power spectrum estimation based on parametric modeling [11, 12]. It has a better frequency resolution and is appropriate to analyze signals of WPT.

Besides feature extraction, pattern recognition is also significant in fault diagnosis. There are some methods like Artificial Neural Network (ANN) [13], Hidden Markov Model (HMM) [14], Support Vector Machine (SVM) etc. Among them, SVM has a better generalization performance comparing with other methods. In addition, it is more appropriate to solve high-dimensional and nonlinear problems. SVM will have premium properties in detection of faults in varying operational conditions because of these traits [4].
Detection of fault in varying operational conditions is still a problem to engineers. This paper takes bearing data of Case Western Reserve University Bearing Data Center on as the object of the research [15] and try to use WPT and AR model spectrum to extract the fault features from its different working condition data, then classifying the faults of different sizes and location [16].

2. Technical background

2.1. Signal processing by wavelet packet decomposition (WPD)

Wavelet transform is a rapidly developing method to non-stationary signals in recent decades. Comparing with traditional short time Fourier transform, it is better some slight benefits in reducing computations when examining specific frequencies and giving a flexible time-frequency domain feature because of the use of variable sized windows. However, it won’t decompose high frequency signals due to its flexibility. Thus, WPD analysis shows up and has a better resolution in high frequency signals. Besides, it draws into the concept of optimal basis selection to select optimal basis function flexibly. The process of WPD is as follows [17].

Assume that \( \varphi_0(x) \) is scaling function and \( \varphi_1(x) \) is wavelet function. The basic wavelet packet functions are defined as

\[
\begin{align*}
\varphi_{2j}(x) &= \sqrt{2} \sum h(k) \varphi_j(2x - k) \\
\varphi_{2j+1}(x) &= \sqrt{2} \sum g(k) \varphi_j(2x - k)
\end{align*}
\]

where \( h(k) \) and \( g(k) \) are the quadrature mirror filters.

For a discrete signal, the decomposition coefficients of wavelet packets can be computed iteratively by

\[
\begin{align*}
E_{2n,j}^k &= \sum_h h(m - 2k)X_{n,j} \\
E_{2n+1,j}^k &= \sum_g g(m - 2k)X_{n,j}
\end{align*}
\]

where \( E_{n,j} \) denotes the wavelet coefficients at the \( j \) level, \( n \) sub-band and \( m \) is the number of the wavelet coefficients.

From equation(1),(2), signals can be decomposed into level \( j \) and the level \( j \) will got \( 2^j \) part of signals by frequency [17]. For example, a wavelet packet decomposition tree of three levels is illustrated in figure1.

![Figure 1. An example of a three-level wavelet packet decomposition tree.](image)

2.2. Auto regressive(AR) model spectrum

Auto regressive model is one of the most widely used and studied parameter models. Auto regressive model spectrum consists of two steps. One is building AR model for time domain signals. The other is calculating the power spectrum by model coefficient [18]. The AR model is defined as

\[
X_j = c + \sum_{i=1}^{p} \varphi_i X_{j-i} + e_j
\]
where $\varphi_1, \ldots, \varphi_p$ are the parameters of the model, $c$ is a constant, and $\epsilon_t$ is white noise.

Equation (3) can be regarded as an input or output equation to a system. Single side spectrum of signals can be calculated by transfer function,

$$S(f) = \frac{\sigma_Z^2}{1 - \sum_{k=1}^{p} \varphi_k e^{-i2\pi f k}}$$

where $\sigma_Z^2 = Var(Z_t)$. Then the energy of vibration in different frequency can be calculated [16].

2.3. Support vector machines with cross-validation

Support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In a prediction problem, a model is usually given a dataset of known data on which training is run, and a dataset of unknown data against which the model is tested. It is always difficult to divide vector sets in low-dimensional space, so they should be mapped to high-dimensional space. It constructs a hyperplane or set of hyperplanes in a high-dimensional or infinite-dimensional space, which can be used for classification. It can be regarded as

$$
\begin{align*}
\min_{\omega,\xi} & \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{n}\xi_i \\
\text{subject to} & \quad y_i(\omega \cdot \Phi(x_i) + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, \forall i
\end{align*}
$$

where $\omega$ is weight coefficient vector of classification hyperplane, $\Phi$ is a linear mapping from low-dimensional space to high-dimensional space, $c$ is penalty parameter, $\xi_i$ is slack variables [19].

However, the complexity will increase a lot and the kernel function will solve the problem. Different kernel functions lead to different SVM algorithms. In machine fault classification, radial basis function (RBF) kernel is a popular kernel function. The RBF kernel on two samples $X_i$ and $X_j$, is defined as

$$K(X_i, X_j) = \exp(-g||X_i - X_j||^2).$$

Then, using Lagrange multiplier method, equation (5) will transform to

$$
\begin{align*}
\min_{a_i} & \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j y_i y_j \exp(-g||X_i - X_j||^2) \\
\text{subject to} & \quad 0 \leq a_i \leq c, \sum_{i=1}^{l} a_i y_i = 0, i = 1, 2, \ldots, l
\end{align*}
$$

where $a_i$ is Lagrange multiplier. The parameters of RBF, $g$, influence the accuracy of classification a lot. The penalty parameter, $c$, is used to control the complexity of the model and approximate error. To get the appropriate parameters, $c$ and $g$, cross-validation will take effect [20].

Cross-validation, sometimes called rotation estimation, is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice [21]. The $k$-fold cross-validation can be used to train. Of the $k$ subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k - 1$ subsamples are used as training data. The cross-validation process is then repeated $k$ times (the folds), with each of the $k$ subsamples used exactly once as the validation data. The $k$ results from the folds can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once [22].
3. Experimental results and analysis

3.1. Feature extraction

The vibration data used at this session have been obtained from the ball bearing test data set of the Case Western Reserve University Bearing Data Center Website. The paper uses the 12kHz sampling frequency data of drive end bearings, which are working 10 seconds in four kinds of conditions, with faults of different size and location. The detailed information is showed in table 1.

| Fault Location | Fault Location |
|----------------|----------------|
| none           | none           |
| Inner Raceway (inch) | Inner Raceway (inch) |
| Outer Raceway (inch) | Outer Raceway (inch) |
| Ball (inch)    | Ball (inch)    |

From the data above, there 40 kinds of data in total. To make best of the data, every kind of data is divided into 40 groups. That means there are 1600 samples of bearing which are working in 0.25 seconds in total.

In generally, wavelet packet transform is used to decompose the signals into 2, 3 or 4 levels. The higher level will lead to more fault information and more computational complexity. So it is necessary to seek balance between them. After decomposing the 1600 samples, we can get each band signals of the samples and use AR model to compute the spectrum energy. Subsequently, the spectrum energy data should be normalized to speed up the convergence rate. Here are 10 kinds of features from random groups after normalization chosen from all the data of every decomposition level in table 2~4 and 40 random samples from every group in figure 2~4. From the example of the data, it is difficult to classify the ten kinds of faults in an ordinary way.

| Table 1. Faults information (fault depth is 0.01 inches). |
|----------------------------------------------------------|
| Fault Location | 0 HP 1797 rpm | 1 HP 1772 rpm | 2 HP 1750 rpm | 3 HP 1730 rpm |
| none           | 0             | 0             | 0             | 0             |
| Inner Raceway (inch) | 0.007 | 0.007 | 0.007 | 0.007 |
| Outer Raceway (inch) | 0.014 | 0.014 | 0.014 | 0.014 |
| Ball (inch)    | 0.021         | 0.021         | 0.021         | 0.021         |
| none           | 0             | 0             | 0             | 0             |
| Inner Raceway (inch) | 0.007 | 0.007 | 0.007 | 0.007 |
| Outer Raceway (inch) | 0.014 | 0.014 | 0.014 | 0.014 |
| Ball (inch)    | 0.021         | 0.021         | 0.021         | 0.021         |

| Table 2. 10 groups of data of 10 kinds of faults in level 2. |
|----------------------------------------------------------|
| none | Ball | Inner Raceway | Outer Raceway |
| $E_{2,0}$ | -0.760 | 0.067 | -0.937 | 0.742 | -0.287 | -0.950 | -0.678 | -0.768 | -0.867 | -0.938 |
| $E_{2,1}$ | -0.999 | 0.272 | -0.964 | -0.078 | -0.716 | -0.943 | 0.495 | -0.852 | -0.963 | -0.981 |
| $E_{2,2}$ | -0.993 | -0.846 | -0.937 | -0.001 | -0.606 | -0.980 | 0.332 | -0.821 | -0.974 | -0.865 |
| $E_{2,3}$ | -1.000 | -0.297 | -0.959 | 0.107 | -0.770 | -0.933 | 0.674 | -0.870 | -0.959 | -0.966 |
### Table 3. 10 groups of data of 10 kinds of faults in level 3.

| none | Ball  | Inner Raceway | Outer Raceway |
|------|-------|---------------|---------------|
| 0    | 0.007 | 0.014         | 0.021         |
| 1.000| -0.842| -0.227        | -0.917        |
| 1.000| -0.620| -0.188        | -0.948        |
| 1.000| -1.000| 0.376         | -0.959        |
| 1.000| -0.961| -0.569        | -0.978        |
| 1.000| -0.999| -0.908        | -0.984        |
| 1.000| -0.989| -0.823        | -0.898        |
| 1.000| -1.000| -0.236        | -0.959        |
| 1.000| -1.000| -0.886        | -0.972        |

### Table 4. 10 groups of data of 10 kinds of faults in level 4.

| none | Ball  | Inner Raceway | Outer Raceway |
|------|-------|---------------|---------------|
| 0    | 0.007 | 0.014         | 0.021         |
| 1.000| 0.433 | 0.769         | 0.877         |
| 1.000| 0.978 | 0.718         | 0.961         |
| 1.000| 0.854 | 0.243         | 0.955         |
| 1.000| 0.529 | 0.221         | 0.966         |
| 1.000| 1.000 | 0.832         | 0.932         |
| 1.000| 1.000 | 0.762         | 0.970         |
| 1.000| 0.971 | 0.056         | 0.936         |
| 1.000| 0.953 | 0.482         | 0.978         |
| 1.000| 0.999 | 0.939         | 0.984         |
| 1.000| 0.999 | 0.799         | 0.988         |
| 1.000| 0.992 | 0.308         | 0.963         |
| 1.000| 0.988 | 0.778         | 0.994         |
| 1.000| 1.000 | 0.823         | 0.928         |
| 1.000| 1.000 | 0.810         | 0.949         |
| 1.000| 1.000 | 0.615         | 0.937         |
| 1.000| 1.000 | 0.678         | 0.977         |
Figure 2. 40 random samples of level 2.

Figure 3. 40 random samples of level 3.

Figure 4. 40 random samples of level 4.
3.2. Fault recognition

In the 40 kinds of features, there are 10 kinds of faults we need to diagnose. To test the effect of SVM, the 1600 groups samples are split into two parts in two ways. One is that half of the features are used for training patterns in every kind of working conditions. The other is that three-quarter of the features are used for training patterns in every kind of working conditions. Then, when using training patterns to train the SVM, k-fold cross-validation is usually used to get a reliable and stable model. According to practice, 10-fold cross-validation is appropriate between complexity and accuracy [22]. At last, we individually predict the faults of 800 test samples and 400 test samples, and calculate the classification rates. The classification rates and running time are shown in table 5 and figure 5, 6.

Table 5. Classification rates and running time in different conditions.

| Decomposition level | 800 training samples, 800 test samples | 1200 training samples, 400 test samples |
|---------------------|---------------------------------------|----------------------------------------|
|                     | 2 levels | 3 levels | 4 levels | 2 levels | 3 levels | 4 levels |
| Classification rates(%) | 98.625 | 99.375 | 100 | 98.75 | 99.5 | 100 |
| Running time(s) | 1814.717 | 2373.86 | 3503.657 | 3695.845 | 4786.401 | 7021.654 |

Figure 5. Comparison of classification rates.  
Figure 6. Comparison of running time.

4. Conclusions

Wavelet packet decomposition can effectively extract fault signals and decrease the interference from varying operational condition signals. The AR model spectrum of the feature signals is ideal to reflect the condition of bearing. And, the SVM model has an excellent performance in classification. Through the result of test, we find that running time increases with decomposition levels increasing. While decomposing from 3 levels to 4 levels, running time will increase about 46.7%, but the classification only increase from 99.5% to 100%. The training samples decrease one third bringing running time decreasing about 50%. In addition, we can also find that the reduction of training samples seldom influences the classification, when training samples account for over half of the total sample. Consequently, it is significant to choose the appropriate decomposition levels and training sample according to requirement and qualification in actual projects.

References

[1] Miao Y, Zhao M, Lin J and Lei Y 2017 Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings Mechanical Systems and Signal Processing 92 173–95
[2] Miao Y, Zhao M and Lin J 2017 Improvement of kurtosis-guided-grams via Gini index for bearing fault feature identification Meas. Sci. Technol. 28 125001
[3] Mcfadden P D and Smith J D 1984 Model for the vibration produced by a single point defect in
a rolling element bearing. Journal of Sound & Vibration 96 69–82

[4] Gryllias K C and Antoniadis I A 2012 A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments Engineering Applications of Artificial Intelligence 25 326–44

[5] Gade S and Herlufsen H 1990 Digital Filter vs FFT techniques for damping measurements Sound & Vibration 96 69–82

[6] Vetterli M and Herley C 1992 Wavelets and filter banks: theory and design (IEEE Press)

[7] Nikolaou N G and Antoniadis I A 2002 Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted morlet wavelets Mechanical Systems & Signal Processing 16 677–94

[8] Gritli Y, Rossi C, Zarri L, Filippetti F, Chatti A, Casadei D and Stefani A 2011 Advanced diagnosis of broken bar fault in induction machines by using Discrete Wavelet Transform under time-varying condition 2011 IEEE International Electric Machines Drives Conference (IEMDC) 2011 IEEE International Electric Machines Drives Conference (IEMDC) pp 424–9

[9] Li C J and Ma J 1997 Wavelet decomposition of vibrations for detection of bearing-localized defects Ndt & E International 30 143–9

[10] Yen G G and Lin K C 2000 Wavelet packet feature extraction for vibration monitoring IEEE Transactions on Industrial Electronics 47 650–67

[11] Wang W and Wong A K 2002 Autoregressive model-based gear fault diagnosis Journal of Vibration & Acoustics 124 172–9

[12] Baille C D and Mathew J 1996 A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings Mechanical Systems & Signal Processing 10 1–17

[13] Jack L B and Nandi A K 2002 Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms Mechanical Systems & Signal Processing 16 373–90

[14] Purushotham V, Narayan S and Prasad S A N 2005 Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition Ndt & E International 38 654–64

[15] Smith W A and Randall R B 2015 Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study Mechanical Systems & Signal Processing 64–65 100–31

[16] Chang Chun, Wang Guo Wei, Mei Jianmin, Zhang Ling Ling and Guo Zheng 2015 Connecting rod bearing fault diagnosis based on the wavelet packet-auto regressive model Spectrum and support vector machine Journal of Military Transportation University 17 40–4

[17] Daubechies I 1992 Ten lectures on wavelets Computers in Physics 6 1671–1671

[18] Storch H V and Zwiers F W 2001 Statistical analysis in climate research (Cambridge University Press)

[19] Shalev-Shwartz S, Singer Y, Srebro N and Cotter A 2007 Pegasos: primal estimated sub-gradient solver for SVM Mathematical Programming 127 3–30

[20] Campbell C 2008 Introduction to Support Vector Machines (Pearson Education/Prentice Hall,)

[21] Geisser S 1993 Predictive inference : an introduction (Chapman & Hall)

[22] Goldstein D R 2004 Analyzing Microarray Gene Expression Data (Wiley-Interscience,)