International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.7 (2015), pp.223-228
http://dx.doi.org/10.14257/ijmue.2015.10.7.23

Research on the AE Signal De-noising Based on K-Means Clustering and the Wavelet Transform

Zhengben Zhang and Chongke Wang

Henan Mechanical and Electrical Engineering College, Xinxiang 453002, China
zhangzhengben@gmail.com

Abstract

The noise in the acoustic emission (AE) signal must be removed to identify the mode of AE signal accurately. The Wavelet threshold de-noising method shows some unique advantages. Based on the threshold selection risky problem, K-means clustering method was used to classify the high-frequency coefficients by the wavelet decomposition to determine the removal threshold for the wavelet coefficients corresponding to the noise, and achieve the de-noising purpose. Hard-threshold method and soft-threshold method were applied to AE signal through the wavelet threshold de-noising. The thresholds generated by K-means clustering approach and the Donoho method improved were respectively used as the threshold for the de-noising of the wavelet coefficients. The experimental results show that the method proposed is superior to the Donoho method improved in the three indicators of signal to noise ratio, root mean square error.

Keywords: *AE signal, K-means clustering, Donoho method, wavelet transform, De-noising*

1. Introduction

AE signal is a non-stationary random signal with low signal-to-noise [1]. The de-noising effect of the traditional linear filter is not ideal [2~4], the wavelet de-noising method of Donoho [5]. has the better de-noising effect [6].This method can almost completely suppress noise and can retain the characteristics of real signals in the greatest degree. Donoho believes that after the wavelet decomposition, the coefficient amplitude distributions of the wavelet of real signal and noise signal are different. The threshold is set to remove the noise wavelet coefficients, and then the wavelet reconstruction is used to achieve the purpose of removing noise.

The threshold selection is a key problem. If the threshold value is too high, the real signal will be lost. If the threshold value is too low, it will influence the de-noising effect. Donoho set up a fixed threshold according to the variance of noise signals and the length of the wavelet decomposition, and did threshold processing for all high frequency coefficients after the wavelet decomposition. The above method is relatively simple, but ignores the characteristic that the noise corresponding wavelet coefficient amplitude increases with decomposition level. Some scholars improved Donoho’s threshold setting method [7]. Donoho’s improved threshold setting method generated the corresponding threshold for each layer of wavelet coefficients in the high frequency to enhance the de-noising effect. However, Donoho’s threshold setting method and its improved method needed know the noise variance, and in the actual calculation, the estimation value was often used, which affected the de-noising effect.

Based on the above the threshold selection problem, the wavelet coefficient classification method was proposed based on the K- means clustering method. The wavelet coefficients of the true signal and the noise signal were divided into two categories. The threshold was determined according to the classification. The improved Donoho method and the K- means clustering method generated respectively the de-

ISSN: 1975-0080 IJMUE

noising threshold for the wavelet threshold de-noising. The experimental results show that this method proposed in this study is superior to the Donoho method improved in the three indicators of signal to noise ratio (*SNR*), root mean square error (*RMSE*) and smoothness

2. The Transform of One-Dimensional Discrete Wavelet

CWT of signal *f* is defined as：

$$(W_\psi f)(a,b)=\langle f,\psi_{a,b}\rangle=|a|^{-\frac{1}{2}}\int_{-\infty}^{+\infty}f(t)\bar{\psi}(\frac{t-b}{a})dt \quad (1)$$

$(W_\psi f)(a,b)$ is the wavelet coefficient, which is the function of the scale *a* and the position *b*. "< >" represents the inner product; $\bar{\psi}$ represents complex conjugate of ψ. Wherein, *a* is the scale parameter; $a\in R$ and $a\neq 0$ represent the stretching and retracting related to the frequency; *b* is the time position parameter [8]. By moving the scale a stretching and retracting and parameter *b* moving, the band pass characteristic of wavelet is used to decompose the signal in the different scale. The translation result can take this group of signal as the window to observe the interesting part. The scale a and offset b are taken as the form of a power series, which is shown as follows:

$$a=a_0^j, b=ka_0^j b_0 \ \ j\in Z \quad (2)$$

Where $a_0\neq 1$is a fixed value, for convenience， $a_0>1$ is always assumed. The corresponding discrete wavelet is presented as:

$$\psi_{j,k}(t)=a_0^{-\frac{j}{2}}\psi(a_0^{-j}t-kb_0) \quad j,k\in Z \quad (3)$$

The discrete wavelet transform coefficient of signal *f* (*T*) is represented as:

$$C_{j,k}=\int_{-\infty}^{+\infty}f(t)\bar{\psi}_{j,k}(t)dt \quad (4)$$

In the actual calculation, it is impossible to calculate wavelet coefficients for all scales and displacement. Furthermore, the actual observed signals are discrete so that the discrete wavelet transform is used in signal processing. In most cases, the scale factor and displacement parameter are discrete according to the power of 2. a_0=2, b_0=1, the scale and offset are discrete for two times and then dyadic wavelet is attained:

$$\psi_{j,k}(t)=2^{-\frac{j}{2}}\psi(2^{-j}t-k) \quad (5)$$

In the *j* layer, signal *f*(*t*) is decomposed into low frequency wavelet coefficient cA_j and the high frequency wavelet coefficients cD_j, which is shown in Figure 1.

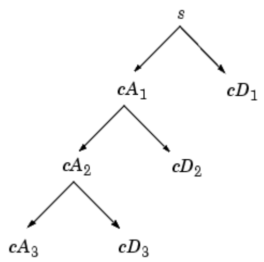

Figure 1. Decomposition Coefficient Structure by One Dimension Dyadic Wavelet

3. The Method of Wavelet Threshold De-noising

After the noise AE signal is transformed through the wavelet, the low frequency coefficients are all reserved. The larger amplitude coefficients in the high-frequency wavelet coefficients are generated by the real signal, while the relatively small coefficients of the amplitude are generated from the noise. If the proper threshold is selected, the high frequency coefficient will be set to zero, while the high frequency coefficient is larger than the threshold to be retained. The processing method can effectively suppress noise of AE signal. Finally, the effective AE signal is obtained after the wavelet reconstruction transformation. This method as the wavelet threshold de-noising is put forward by Donoho. The threshold generation method of Donoho includes the hard threshold and soft threshold methods.

The hard threshold：$$\hat{C}_{j,k} = \begin{cases} C_{j,k}, \left|C_{j,k}\right| \geq T \\ 0, \left|C_{j,k}\right| < T \end{cases} \quad (6)$$

The soft threshold：$$\hat{C}_{j,k} = \begin{cases} \mathrm{sgn}(C_{j,k})(\left|C_{j,k}\right| - T), \left|C_{j,k}\right| \geq T \\ 0, \left|C_{j,k}\right| < T \end{cases} \quad (7)$$

Donoho put forward the universal threshold:

$$T = \sigma\sqrt{2\ln N} \quad (8)$$

Hard threshold method is relatively simple, but when it eliminates the noise, it may remove a part of useful signals. The soft threshold is used to remove noise through contracting the amplitude of the wavelet coefficients, to reduce the risk of the threshold, and retain the characteristics of the original signal as much as possible. Furthermore, the signal is smoother by the soft threshold de-noising method.

4. The Wavelet De-noising Threshold Generation Based on K-Means Clustering Method

K-means clustering method is a typical kind of the clustering algorithm based on distance [9]. Its input $X=\{x_1,x_2,\ldots, x_n\}$, and classification number is k. The output is k data type C_j, $j=1,2,\ldots,k$. The main steps are shown as follows:

1. Randomly in the dataset, k initial cluster centers are specified ($m_1,m_2,\ldots, m_k$).

2.For the data element x_i,it is necessary to calculate the distance $d(x_i,m_j),i=1,2,\ldots,n,j=1,2,\ldots,k$, between k initial cluster centers, which usually uses Euclidean distance [10]. If it can satisfy

$$d(x_i, m_p) = \min\{d(x_i, m_j), j = 1, 2, ..., k\} \quad (9)$$

Then $x_i \in C_p$.

3. The k new clustering centers are updated. All mean values are regarded as the new cluster centers. The error square and function are treated as the convergence function:

$$E = \sum_{j=1}^{K} \sum_{x \in C_j} \left\| x - m_j \right\|^2 \quad (10)$$

4. If the E value is the convergence, the final clustering is output. Otherwise, step 2 is turned.

The high frequency wavelet coefficients magnitude cD_j is regarded as the input data set, and clustering number k is 2, which is corresponding to the wavelet coefficient amplitude of the true signal and the amplitude of the noise wavelet coefficients respectively. K-means clustering method is employed to do the iterative classification to the high frequency coefficients of each scale. Then, according to the conditions of convergence, the critical value between classifications is determined, which is the threshold de-noising of the of the wavelet coefficients.

5. The Results and Analysis of Experiments

In Figure 2, continuous AE signal is based on Matlab R2010a as the experimental platform. One-dimensional discrete wavelet decomposition is for the AE signal, and db2 wavelet is selected as the wavelet base with decomposition layer 4, which is shown in Figure 3.

After the wavelet decomposition, the de-noising threshold setting of the high frequency coefficients of each level adopts respectively the improved method of Donoho and the proposed method based on K-means clustering. The K-means clustering method is used to classify the amplitude of high frequency coefficient is shown in Figure 4.

In Table 1, two methods in different high frequency coefficients generate the wavelet de-noising thresholds.

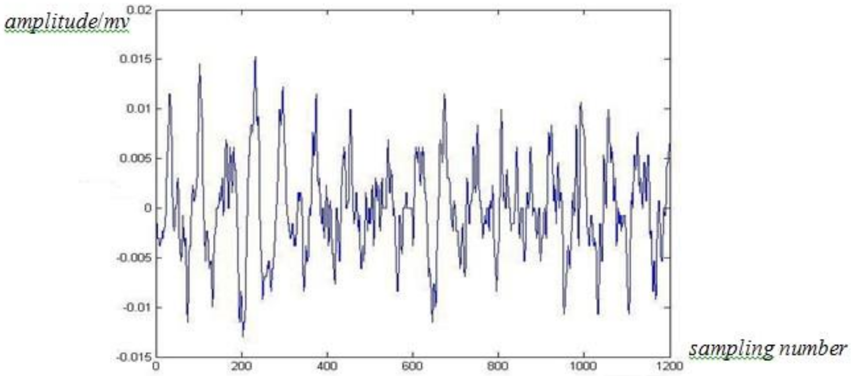

Figure 2. AE Signal n Time Domain

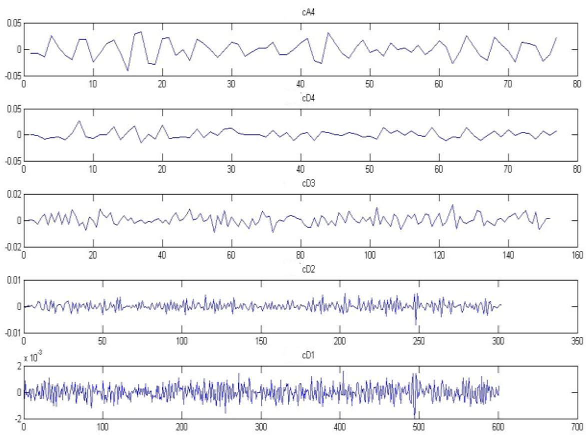

Figure 3. Four Layer Decomposition Structure by db2 Wavelet

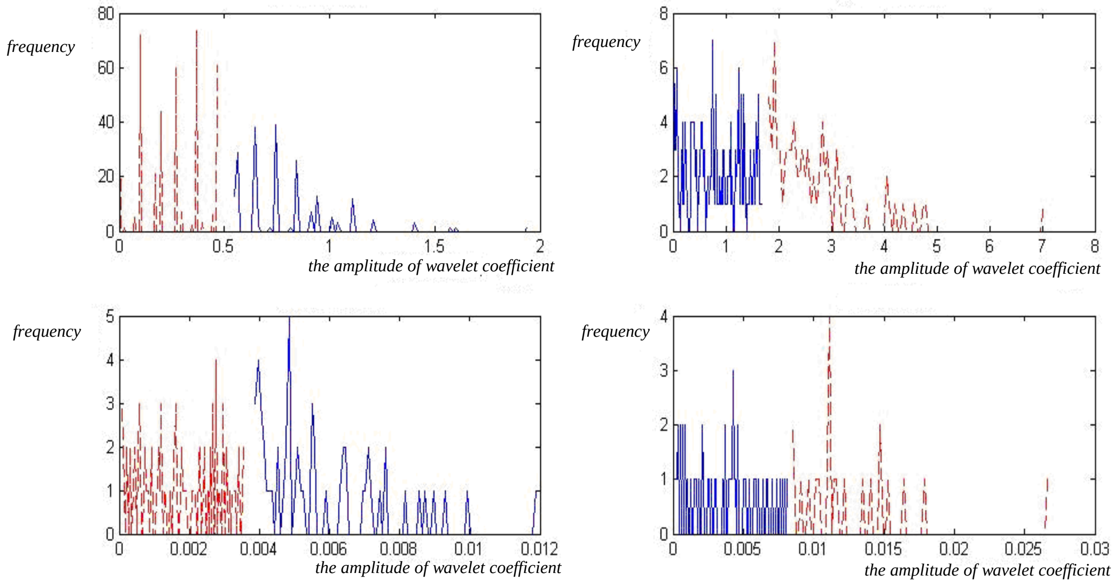

Figure 4. Classify for High Frequency Coefficients by wavelet Decomposition

Table 1. Threshold on High Frequency Coefficients by Two Different Methods

Threshold generation method	cD_1	cD_2	cD_3	cD_4
Donoho's improved method	0.0022	0.0061	0.0014	0.0161
K-means clustering method	0.0003	0.0015	0.0023	0.0081

Note: cD_j (j=1-4) is the high frequency coefficients of the j layer

In Table 1, Donoho's improved method and K-means clustering method generate the thresholds. SNR and RMSE are employed to compare the noise reduction performance, which is shown in Table 2. RMSE embodies the differences between the original signals and denoising signals. The smaller the threshold is; the better the de-noising effect is. SNR refers to the ratio between the original signal energy and the noise energy [10]. Generally, the higher the SNR is; the better the effect of noise filtering is.

Table 2. Performance Comparison on De-noise of AE Signal

Denoisng method	Threshold by Donoho's improved method		Threshold by K-means clustering method	
	SNR	*RMSE*	*SNR*	*RMSE*
Hard threshold method	7.2312	0.0105	9.1211	0.0031
Soft threshold method	4.8304	0.0120	12.2823	0.0011

Based on the experimental results, the wavelet threshold method is used for AE signal in the signal de-noising effect. Generating threshold by the K-means clustering method is better than that by the improved Donoho method, and soft threshold method is better than hard threshold method in the de-noising effect.

6. Conclusion

The wavelet threshold de-noising method proposed by Donoho has a unique advantage of AE signal. In the wavelet threshold de-noising method, the key problem is in threshold settings. If the threshold selected is too large, the real signal will be lost. If the threshold selected is too small, it will influence the de-noising effect. Improved Donoho's threshold setting method need know the noise variance. In the actual calculation, estimation value is often used, which will influence the de-noising effect. The wavelet coefficient classification method was proposed based on the K-means clustering method. The wavelet coefficients of the true signal and the noise signal were divided into two categories. The threshold was determined according to the classification. The improved Donoho method and the K-mean clustering method generated respectively the de-noising threshold for the wavelet threshold de-noising. The experimental results show that this method proposed in this study is superior to the Donoho's improved method in indicators of signal to noise ratio, and root mean square error.

References

[1] M. Y. Li and Z. D. Shang, "Acoustic emission detection and signal processing," Science Press, Beijing **(2010)**.
[2] Y. Q. Zhang, W. F. Niu and W. Tan, "The characteristics Analysis based on wavelet transform of acoustic emission signal of simulation," Nondestructive testing journal, vol. 33 no. 1, **(2011)**, pp. 16-17.
[3] G. H. Li and S. F. Liu, "The analysis technology and development of AE signal," The 10th National AE conference proceedings, Daqing, China, **(2004)**.
[4] J. L. Li, Y. K. Ma and Y. Z. Dong, "The study and analysis of AE detection using artificial neural network noise," Nondestructive testing, vol. 21 no. 9, **(1999)**, pp. 399.
[5] D. L. Donoho and I. M. Johnstone, "Wavelet Shrinkage: A symptopia," Journal of the Royal Statistical Society, vol. 57, **(1995)**, pp. 301- 369.
[6] R. G. Yang, "Wavelet de-noising using principal component analysis," Expert Systems with Applications, vol. 38, **(2011)**, pp. 1073- 1076.
[7] W. Q. Zhang and G. X. Song, "De-noising in wavelet domain signal based on the new threshold function," Journal of Xi'an Electronic and Science University, vol. 31 no. 2, **(2004)**, pp. 296-299
[8] I. Daubechies, "The wavelet transform: Time- frequency localization and signal analysis," IEEE 2000, vol. 36, **(2010)**, pp.961- 1005.
[9] S. Sambasivam and N. Theodosopoulos, "Advanced data clustering methods of mining web documents," Issues in Informing Science and Information Technology, vol. 8 no. 3, **(2006)**, pp.563-579.
[10] K. Tao and J. J. Zhu, "A comparative study on evaluation method of wavelet noise quality," Geodesy and geodynamics, vol. 32 no. 2, **(2012)**, pp.129-130.

Authors

Zhengben Zhang, received his Master Degree from Wuhan University Of Technology in 2010. He is now a experimentalist of Henan Mechanical and Electrical Engineering College in China. His research interests include software engineering, computer graphics. He has published more than 6 papers on journals and conferences.

Chongke Wang, received his Master Degree from Wuhan University Of Technology in 2009. He is now an experimentalist of Henan Mechanical and Electrical Engineering College in China. His research interests include software engineering, data mining. He has published more than 5 papers on journals and conferences.

