

Article



A Method of Barkhausen Noise Feature Extraction Based on an Adaptive Threshold

Cheng Hang ^{1,2,*}, Wenbo Liu ^{1,2} and Ping Wang ^{1,2}

- ¹ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211006, Jiangsu, China
- ² Nondestructive Detection and Monitoring Technology for High Speed Transportation Facilities, Key Laboratory of Ministry of Industry and Information Technology, Nanjing 211006, Jiangsu, China
- * Correspondence: hangcheng@nuaa.edu.cn

Received: 31 May 2019; Accepted: 22 July 2019; Published: 24 July 2019



Abstract: This paper reports on a new feature extraction method for detection of applied stress using magnetic Barkhausen noise (MBN). Some previous methods for extracting MBN features need to choose a suitable threshold so that these features can have good linearity and low dispersion, such as pulse count and full width at 25, 50 and 75% of the maximum amplitude. A new approach has been proposed for selecting the appropriate threshold for MBN features adaptively using a genetic algorithm (GA). The criterion for selecting the threshold is the lowest standard deviation of features and new proposed 'overlap' of features. In order to verify the effectiveness of the adaptive pulse count feature for stress detection, different modelling techniques are compared, including multivariable linear regression (MLR) and multilayer perceptron (MLP). The results obtained have proven that adaptive threshold features can effectively distinguish between different stress conditions compared with traditional MBN features.

Keywords: magnetic Barkhausen noise (MBN); adaptive feature extraction; threshold; genetic algorithm

1. Introduction

The magnetic Barkhausen noise (MBN) as a non-destructive evaluation method is mainly used to detect stress of ferromagnetic materials [1–3]. Feature extraction of MBN signals is important for stress detection of ferromagnetic materials. In order to preserve more information of the MBN signals, data of the collected MBN signals are relatively large. If a stress model is established directly with the original MBN signals, a large amount of redundant and irrelevant information will lead to a prolonged time spent on training the model and using the model for stress detection, and will reduce measurement accuracy. Therefore, how to extract MBN features is critical to the results of stress detection.

However, previous methods for extracting MBN features require the choice of a suitable threshold so that these features can have good linearity and low dispersion. Figure 1 shows two novel features of MBN signals, namely 'pulse count' and 'event'. It was found that new features had a good linear correlation with residual stress [1]. However, the linearity and dispersion of the pulse count feature are closely related to the choice of threshold.



Figure 1. Illustration of 'pulse count' and 'event' of MBN signals.

Similarly, MBN features extracted by the commercial non-destructive testing instrument 3MA are full width at 25, 50 and 75% of the maximum amplitude [2,3], which is shown in Figure 2. These three thresholds are not necessarily the best choice for stress detection or other parameter measurements.



Figure 2. Illustration of full width at 25, 50 and 75% of the maximum amplitude.

In this paper, a new approach has been proposed for selecting the appropriate threshold for MBN features adaptively. The criterion for selecting the threshold is the lowest standard deviation and 'overlap' of features. We use a genetic algorithm (GA), a global optimization algorithm, which is adopted to find the optimal threshold. The adaptive pulse count feature and adaptive full width feature have better linearity and stability than features of previous fixed threshold methods [1,2]. Modeling the traditional feature root mean square (RMS), mean value, energy, entropy and new adaptive pulse count feature by multivariable linear regression (MLR) and multilayer perceptron (MLP), the results obtained have proven that adaptive threshold features can effectively improve the accuracy of stress detection.

The rest of the paper is organized as follows. In Section 2, the details of the sample, MBN measurement systems and process of adaptive feature extraction are illustrated. The results of adaptive threshold features selected by genetic algorithm and 'overlap' criterion are given in Section 3.1. The multi-threshold adaptive extraction method is introduced in Section 3.2. Adaptive pulse count feature and traditional MBN features are compared in Section 3.3. In order to verify the effectiveness of adaptive pulse count feature for stress detection, different modelling techniques are compared in Section 3.4. In Section 4, the summary and conclusion are provided.

2. Experiment

Based on previous work [4], specimens used in the experiment are made of Q235 steel, which is a standard material in industrial manufacturing. A four-point bending platform is used to generate applied stress. The applied tensile and compressive stress is limited in the range of 0–150 MPa to avoid plastic deformation.

The block diagram of the MBN measurement system is shown in Figure 3. To obtain MBN signals, the function generator provides low frequency excitation (triangle wave at 5 Hz) which is amplified by a bi-polar power amplifier (LPA05B). The MBN signals are detected using a pick-up air-coil (with 3000 turns of 0.07 mm diameter wire), which are amplified (30 dB gain), band pass filtered (2–40 kHz), and sampled at 200 kHz frequency by A/D converter (14-bit A/D resolution, DAQ2010). In addition, the excitation voltage is acquired using a second channel synchronously. An example of MBN signal and excitation signal are shown in Figure 4.



Figure 3. Magnetic Barkhausen noise (MBN) measurement systems.



Figure 4. Example of an MBN signal and excitation signal. The red curve is the excitation signal and the blue one is the MBN signal.

In the experiment, five cycles of MBN signals are sampled under the same stress condition, so the number of MBN signals is 10. Figure 5 shows pulse count as a function of stress when threshold = 0.05, the negative stress represents compressive stress, and the positive stress represents tensile stress.



Figure 5. Example of pulse count as a function of stress when threshold = 0.05.

MBN features are normalized for the reason that it is feasible to compare the features of different thresholds. The range of normalized features is between [0,1]. The normalized equation is as follows:

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_i is the *i*th feature of MBN under different stress conditions, x_{min} is the minimum value of features, x_{max} is the maximum value of features, y_i is the result of normalization. Figure 6 shows the difference of pulse count before and after normalization with different thresholds as a function of stress. The x-axis represents different thresholds from 0.050 to 0.032. The pulse count values are changed from -150 MPa to 150 MPa with each threshold, which is equivalent to Figure 5 under different stress conditions.

The number of MBN signals under the same stress condition is 10 (five cycles) and the number of stress conditions is 11 (-150 MPa-150 MPa). The standard deviation of the pulse count feature of 10 MBN signals (under the same stress) can be calculated. So sum of the standard deviations is defined by the sum of the standard deviations in the 11 stress conditions.



(a) Diagram of unnormalized pulse count with different thresholds as a function of stress.



(b) Diagram of normalized pulse count with different thresholds as a function of stress.

Figure 6. Diagram of pulse count before and after normalization with different thresholds as a function of stress.

A genetic algorithm is an optimization technique with the basic idea derived from biological evolution [5–7]. The algorithm is based on the population of chromosomes that achieves a global optimal solution by selection, crossover and mutation. The parameters of the genetic algorithm in the paper are set as follows: Population size (Npop = 40), the number of generations (Ngen = 300) and the probabilities of crossover (pc = 0.8) and mutation (pm = 0.05), because only one optimal threshold

needs to be found in this article. *N*pop determines the search number of thresholds, which is set to 40. *N*gen determines the number of iterations of algorithm. *pc* determines the crossover probability of binary chromosome. Two-point crossover is used in the paper, which picks two crossover points from the parent chromosomes randomly. *pm* determines the probability of altering one or more gene values in a chromosome [8].

We use this genetic algorithm to choose an appropriate threshold for MBN signal features. There are two criteria for evaluating features including the sum of standard deviation and new proposed 'overlap'. As shown in Figure 7, the steps are as follows.

- 1. Extract MBN signal features based on threshold, such as pulse count feature or full width feature.
- 2. Normalize features to [0,1] according to Equation (1) so that features can be evaluated by two criterions.
- 3. Calculate the sum of standard deviation or 'overlap' of features.
- 4. Threshold is adjusted by genetic algorithm to minimize the sum of the standard deviation and 'overlap'. The criteria for selecting the optimal threshold is when 'overlap' is 0 and the sum of standard deviation is the lowest.



Figure 7. Flowchart of adaptive threshold selection based on genetic algorithm.

3. Results

3.1. Adaptive Pulse Count Feature

As shown in Figure 8, the pulse count with different thresholds as a function of stress and standard deviation of every feature are shown. With Table 1 and Figure 9, in the case of large stress (130 MPa and 150 MPa tensile and compressive stress), thresholds of 0.038, 0.036 and 0.034 are difficult to distinguish stress conditions, because the variation range of feature values has overlapping parts. Threshold of 0.024 can distinguish different stress conditions well, whose variation range has little overlapping parts, and the sum of standard deviations is lower.



Figure 8. Diagram of the standard deviation of the pulse count feature with different thresholds.



Figure 9. Distinction of pulse count feature with different thresholds.

Table 1. Sum of the standard deviation of the pulse count feature with different thresholds.

Threshold	Sum of Standard Deviation	Threshold	Sum of Standard Deviation
0.020	0.1794	0.030	0.1652
0.022	0.1675	0.032	0.1685
0.024	0.1600	0.034	0.1753
0.026	0.1585	0.036	0.1781
0.028	0.1554	0.038	0.1797

However, it is difficult to manually find the appropriate threshold, so a genetic algorithm is used in this paper. After optimization by GA, when the threshold is 0.0249, the sum of standard deviation is the lowest, which is 0.1501. The sum of standard deviation varies with genetic generation as shown in Figure 10; it can be found that the genetic algorithm can find the optimal threshold in about 70 generations.



Figure 10. The sum of standard deviation varies with genetic generation.

Since the sum of standard deviation can only reflect the fluctuation range of the whole feature, it cannot reflect the distinguishing degree of features for different stress conditions. The paper proposes a new criterion 'overlap' for judging the distinguishing degree of features. Figure 11 depicts the calculation method of 'overlap'. Assume *stress2* is greater than *stress1*(tensile stress) or *stress2* is less than *stress1*(compressive stress), 'overlap' can be divided into three cases in Figure 11 and be calculated by Equation (2), where $S1_{up}$ and $S1_{low}$ are the upper and lower limits of the feature change under the condition of *stress1*, $S2_{up}$ and $S2_{low}$ are the upper and lower limits of the feature change under the

condition of *stress2*. The lower the 'overlap', the greater the ability of this feature to distinguish from different stress conditions.



Figure 11. Illustration of criterion 'overlap'.

Table 2 shows the sum of overlap of the pulse count feature with different thresholds. When the threshold is 0.020, the sum of overlap is equal to 0, which indicates that this threshold is more suitable for stress detection than other ten thresholds. If there is a case where 'overlap' of multiple pulse count features with different thresholds is 0, the feature with the lowest sum of standard deviation can be selected as the most optimal threshold. After optimization by the genetic algorithm, when the threshold is 0.0230, 'overlap' is 0 and sum of standard deviation is 0.1633, which is the lowest.

Threshold	Sum of Overlap	Threshold	Sum of Overlap
0.020	0	0.030	0.02169
0.022	0.00375	0.032	0.02587
0.024	0.00374	0.034	0.02708
0.026	0.01921	0.036	1.00614
0.028	0.01388	0.038	1.00876

Table 2. Sum of overlap of pulse count feature with different thresholds.

3.2. Multiple Thresholds

As shown in Figure 8, the standard deviation of tensile stress (30 MPa–150 MPa) when threshold is equal to 0.020 is less than threshold equal to 0.024, but the result of the sum of the standard deviations is the opposite. Therefore, for tensile stress and compressive stress, different thresholds can be chosen to obtain more stable features, using the genetic algorithm to optimize the pulse count thresholds of MBN signals under tensile and compressive stress. As shown in Figure 12, for tensile stress, the threshold is chosen to be 0.034, and for compressive stress, the threshold is chosen to be 0.025, and the sum of standard deviations is the lowest.



Figure 12. Diagram of multi-threshold pulse count feature as a function of stress (threshold = 0.034 for tensile stress, threshold = 0.025 for compressive stress).

3.3. Comparison with Traditional MBN Features

Traditional MBN features such as root mean square (RMS) and MBN energy are often used for stress detection [9,10].

Energy feature describes the strength of the signal, and the calculation formula is the sum of the square of the amplitude of each point signal. The calculation formula is as follows:

$$Energy_{MBN} = (x_1^2 + x_2^2 + \dots + x_n^2)$$
(3)

where $x_1 x_2 \dots x_n$ is the amplitude of MBN discrete signal, and *n* is the number of MBN signal points.

Normalized RMS, energy and the adaptive pulse count feature are compared in Figure 13. The sum of standard deviations of different features is shown in Table 3. It can be found that the adaptive pulse count feature proposed in this paper has a lower standard deviation. Adaptive pulse count has low dispersion and better stability.



Figure 13. The figure compares normalized root mean square (RMS), energy and the adaptive pulse count feature under different stress conditions.

Table 3. The sum of standard deviation and sum of 'overlap' of different features.

Feature	Sum of Standard Deviation	Sum of Overlap
RMS	0.1720	0.0121
Energy	0.1987	0.0139
Adaptive Pulse Count	0.1633	0

3.4. Comparison of Modeling Results

The MBN features used in this paper to establish a stress prediction model are: RMS (x_1), mean value (x_2), energy (x_3), entropy (x_4) and adaptive pulse count (x_5).

Mean value describes the magnitude of the DC (Direct Current) component in MBN signals. The entropy feature is a measure of the uncertainty of a random variable [11]. For a discrete signal, if the possible value is $\{x_1, x_2, ..., x_n\}$, the corresponding probability is $\{P(x_1), P(x_2), ..., P(x_n)\}$, and entropy can be calculated as follows:

$$Entropy_{MBN} = -\sum_{i} P(x_i) \log_b P(x_i).$$
(4)

When b = 2, the unit of entropy is bit; when b = e, the unit of entropy is nat; when b = 10, the unit of entropy is Hart. Entropy can be used to measure the uncertainty of MBN signals. The larger the uncertainty of MBN signal, the larger entropy feature.

K-Fold Cross-Validation is used to verify the model [12]. The steps are as follows: (1) Divide the data set into K; (2) Each time one of them is used as a test set, the other K - 1 are used as a training set to train the model; (3) Mean square error (MSE) of the model on the test set is calculated; (4) Repeat the above K times, and average the MSE to obtain the evaluation criteria of the model (Root Mean Square Error, $RMSE_{CV}$). 10-Fold Cross-Validation is used in the paper.

$$RMSE_{CV} = \sqrt{\frac{1}{K} \sum_{i=1}^{K} MSE_i}.$$
(5)

In order to verify the effectiveness of adaptive pulse count feature for stress detection, different modelling techniques are compared, including multivariable linear regression (MLR) and multilayer perceptron (MLP). Coefficients of MLR are solved by the least square method. MLP is trained using backpropagation (BP). The square error is used as the loss function. The advantage of MLP is that the model learned is nonlinear.

The *RMSE*_{*CV*} and coefficient of determination (\mathbb{R}^2) of the comparison experiment are shown in the Table 4. $x_{\text{th}=0.1}$ and $x_{\text{th}=0.01}$ are features of pulse count extracted when threshold is 0.1 and 0.01 respectively.

Line	Method	Features	RMSE _{CV}	R ²
1	MLR	$x_1 x_2 x_3 x_4$	14.084	0.978
2	MLR	$x_1 x_2 x_3 x_4 x_5$	13.307	0.980
3	MLR	$x_1 x_2 x_3 x_4 x_{\text{th}=0.1}$	13.395	0.980
4	MLR	$x_1 x_2 x_3 x_4 x_{\text{th}=0.01}$	14.672	0.978
5	MLP (BP)	$x_1 x_2 x_3 x_4$	6.468	0.995
6	MLP (BP)	$x_1 x_2 x_3 x_4 x_5$	6.002	0.996
7	MLP (BP)	$x_1 x_2 x_3 x_4 x_{\text{th}=0.1}$	6.251	0.995
8	MLP (BP)	$x_1 x_2 x_3 x_4 x_{\text{th}=0.01}$	8.324	0.992

Table 4. Comparison of different modeling methods. Multivariable linear regression (MLR) and multilayer perceptron (MLP).

(1) There are two hidden layers in the MLP method, but different numbers of hidden layer nodes have the same effect on the results. The number of hidden nodes in the Table 4 is 10,10.

(2) By comparing the first and second lines, it can be found that adding the adaptive pulse count feature (x_5) can improve the accuracy of stress detection.

(3) By comparing the second, third and fourth lines, pulse count features of other thresholds $(x_{\text{th}=0.1}, x_{\text{th}=0.01})$ reduce the prediction accuracy of the model. If the threshold is not chosen properly, the prediction accuracy is worse than not adding the pulse count feature.

(4) The results of the MLR and MLP models show that the adaptive pulse count feature (x_5) has a lower RMSE, which contributes to stress detection.

The stress prediction model established by MLR method is as follows:

 $stress = -128.13 - 1472.04x_1 - 317.04x_2 + 1197.73x_3 + 541.36x_4 + 339.57x_5$

 $stress = -114.63 - 1208.64x_1 - 540.88x_2 + 1032.73x_3 + 983.47x_4 + 12.97x_{th=0.01}$

The regression coefficient of x_5 is greater than $x_{th=0.01}$. When examining the contribution of independent variables to dependent variables, it is often not enough to calculate the regression coefficients, and the correlations need to be considered together. The Pearson correlation coefficient for each feature is as follows: x_1 : 0.9635; x_2 : 0.9606; x_3 : 0.9548; x_4 : 0.9688; x_5 : 0.9642; $x_{th=0.01}$: 0.9512. It can be found that the contribution of adaptive pulse count feature to stress is greater than the features of pulse count extracted when threshold is 0.01.

4. Conclusions

In this paper, appropriate thresholds are selected adaptively by standard deviation criterion and 'overlap' with a genetic algorithm for MBN features. The criteria for selecting the optimal threshold is when 'overlap' is 0 and the sum of standard deviation is the lowest.

- 1. The results obtained have proven that adaptive threshold features can effectively distinguish between different stress conditions. Adaptive pulse count has good stability and repeatability.
- 2. Compared with a single threshold, different thresholds are selected for tensile stress and compressive stress respectively, the obtained features have better linearity and stability.
- 3. By comparing different modeling methods including MLR and MLP, the results show that adding the adaptive pulse count feature to stress prediction model will result in a lower RMSE and can improve the accuracy of stress prediction.
- 4. The new method proposed in the paper can be used for other feature extraction situations where thresholds need be selected. For an incremental permeability (IP) signal, curve width at the certain percentage of μMAX can be selected optimally [3]. For multi-frequency eddy current testing, the appropriate excitation frequency also can be chosen optimally.

Author Contributions: Writing: C.H.; Formal Analysis: C.H.; Software: C.H.; Validation: C.H.; Visualization: C.H.; Funding acquisition: W.L. and P.W.; Project administration: W.L. and P.W.; Data curation: P.W.; Supervision: W.L. All authors were involved in experimental investigations.

Funding: This work was supported in part by Aviation science funds under Grant 2015ZF52067, in part by Special scientific instrument development of Ministry of science and technology of China under Grant 2016YFF0103702, in part by National Natural Science Foundation of China under Grant 61527803, in part by National key research and development program of Ministry of science and technology of China under Grant 2016YFB1100205, in part by National quality foundation key research and development plan of common research and application under Grant 2017YFF0209700, and in part by Graduate innovation center fund NUAA under Grant KFJJ20180317.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Vashista, M.; Paul, S. Novel processing of Barkhausen noise signal for assessment of residual stress in surface ground components exhibiting poor magnetic response. *J. Magn. Magn. Mater.* 2011, 323, 2579–2584. [CrossRef]
- 2. Altpeter, I.; Becker, R.; Dobmann, G.; Kern, R.; Theiner, W.; Yashan, A. Robust solutions of inverse problems in electromagnetic non-destructive evaluation. *Inverse Probl.* **2002**, *18*, 1907–1921. [CrossRef]
- 3. Wolter, B.; Gabi, Y.; Conrad, C. Nondestructive Testing with 3MA—An Overview of Principles and Applications. *Appl. Sci.* **2019**, *9*, 1068. [CrossRef]

- 4. Ding, S.; Tian, G.; Dobmann, G.; Wang, P. Analysis of domain wall dynamics based on skewness of magnetic Barkhausen noise for applied stress determination. *J. Magn. Magn. Mater.* **2017**, *421*, 225–229. [CrossRef]
- Sorsa, A.; Leiviskä, K.; Santa-aho, S.; Vippola, M.; Lepistö, T. An Efficient Procedure for Identifying the Prediction Model Between Residual Stress and Barkhausen Noise. *J. Nondestruct. Eval.* 2013, 32, 341–349. [CrossRef]
- Sorsa, A.; Ruusunen, M.; Leiviskä, K.; Santa-Aho, S.; Vippola, M.; Lepistö, T. An Attempt to Find an Empirical Model between Barkhausen Noise and Stress. *Mater. Sci. Forum* 2013, 768–769, 209–216. [CrossRef]
- 7. Michalewicz, Z. *Genetic Algorithms* + *Data Structures* = *Evolution Programs*; Springer Science & Business Media: Berlin, Germany, 2013.
- Gwiazda, T.D. Genetic Algorithms Reference Vol.1 Crossover for Single-Objective Numerical Optimization Problems; Tomaszgwiazda e-Books: Lomianki, Poland, 2006; Available online: https://books.google.com/books?id= wV68jqT-YJ4C (accessed on 24 May 2019).
- Sorsa, A.; Leiviskä, K.; Santa-Aho, S.; Lepistö, T. A data-based modelling scheme for estimating residual stress from Barkhausen noise measurements. *Insight Non-Destructive Test. Cond. Monit.* 2012, 54, 278–283. [CrossRef]
- 10. Stefanita, C.G. Plastic deformation effects on magnetic Barkhausen noise. *AIP Conf. Proc.* **2000**, 509, 1541–1548. [CrossRef]
- 11. Cover, T.M.; Thomas, J.A. Elements of Information Theory; John Wiley & Sons: Hoboken, NJ, USA, 2012.
- 12. Mclachlan, G.J.; Do, K.A.; Ambroise, C. Analyzing Microarray Gene Expression Data. In *Analyzing Microarray Gene Expression Data*; John Wiley & Sons: Hoboken, NJ, USA, 2014.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).