Fault causes identification for transmission lines based on HHT and PNN

Wen Du¹, Renfei Che¹,a, Hao Zhang¹, Xiaocheng Meng¹,² and Juntao He¹,³

¹Key Laboratory of Power System Intelligent Dispatch and Control of the Ministry of Education, Shandong University, Jinan; 250061, China
²State Grid Jiangsu Electric Power Company Technician Training Center, Suzhou; 215004, China
³Zhengzhou Power Supply Company, Zhengzhou; 450006, China

Abstract. In this paper, Hilbert-Huang transform (HHT) is introduced into the application of fault cause identification for transmission lines. This method is composed of empirical mode decomposition (EMD) and Hilbert transform. It can analyze the characteristics of signal frequency quantitatively and accurately[1]. By analyzing the information of four common faults from fault recorder, the Hilbert marginal spectrum and the Hilbert time-frequency spectrum can be considered as the feature of different fault causes which can be classified by a probability neural network (PNN). The results show that the method can effectively identify the cause of the fault and has a high accuracy.

1. Introduction
Transmission line faults is one of the frequent grid accidents, which is harmful to the operation of the power system and the reliability of the power supply. Affected by complicated geographical and climatic environment, transmission lines are vulnerable to various forms of damage such as lightning strikes, birds and beasts, pollution, and external forces [2,3]. Therefore, it’s necessary to study the characteristics of various faults causes and distinguish them, which can provide support for professionals to find and eliminate faults timely.

At present, fault recorders have been widely used in power grids of different voltage levels, which can effectively record fault information. In this paper, the HHT method is used to analyze the recorded information and the PNN is used as a classifier to construct a fault cause identification system, which achieves the expected goal.

2. The Hilbert-Huang transform
The Hilbert-Huang transform is a new method for analyzing nonlinear, non-stationary signals and has been widely used in engineering. HHT is capable of accurate time-frequency and spectral analysis with high resolution. The Hilbert-Huang transform consists of two steps. First, empirical mode decomposition (EMD) is performed to obtain a finite number of intrinsic mode functions (IMFs). Then, the Hilbert transform is applied to these intrinsic mode functions to obtain the Hilbert time-frequency spectrum and Hilbert marginal spectrum.
2.1 The empirical mode decomposition

The intrinsic mode function calculated by empirical mode decomposition must meet the following two conditions:

① The number of extreme points and zero-crossings must be equal or differ at most by 1.

② At any instant, the mean value of the envelope defined by the local maxima and the local maxima is zero [1].

Considering a given signal $S(t)$, the symmetric envelopes can be obtained respectively by the local maxima and minima of the signal data series. Then by calculating the margin $M_1(t)$ of the upper and lower envelopes a new signal $H_1(t)$ is obtained by

$$H_1(t) = S(t) - M_1(t)$$ (1)

If $H_1(t)$ does not satisfy the above two conditions, then $H_1(t)$ is regarded as a new signal, and the margin $M_{11}(t)$ is again calculated to obtain the $H_{11}(t)$

$$H_{11}(t) = H_1(t) - M_{11}(t)$$ (2)

This cycle will stop when a $H_{1K}(t)$ can satisfy the IMF conditions. Then the first intrinsic mode function is

$$I_1(t) = H_{1K}(t)$$ (3)

The residual signal $S_1(t)$ is obtained by

$$S_1(t) = S(t) - I_1(t)$$ (4)

By repeating the above process, a series of intrinsic mode functions can be obtained, and the original signal can be described as

$$S(t) = \sum_{j=1}^{N} I_j(t) + S_N(t)$$ (5)

$N$ is the number of intrinsic mode functions, $I_j(t)$ is the $j$th intrinsic mode component, and $S_N(t)$ is the residue.

2.2 The Hilbert transform

The Hilbert transform is performed on the intrinsic mode functions obtained by EMD, so that the instantaneous frequency of the original signal has physical meaning, and then the Hilbert time-frequency spectrum and the Hilbert marginal spectrum can be obtained. For any intrinsic mode function $I_j(t)$, its Hilbert transform is as follows

$$v_j(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{I_j(\tau)}{t-\tau} \, d\tau$$ (6)

Take $I_j(t)$ as the real part and $v_j(t)$ as the imaginary part to construct a signal

$$z_j(t) = I_j(t) + iv_j(t)$$ (7)

Instantaneous amplitude $A_j(t)$ and instantaneous frequency $f_j(t)$ are obtained by

$$\theta_j(t) = \tan^{-1} \frac{v_j(t)}{I_j(t)}$$ (8)

$$f_j(t) = \frac{1}{2\pi} \frac{d\theta_j(t)}{dt}$$ (9)

3. Fault current analysis and feature extraction by HHT

In this paper, the zero-sequence current is taken as the research object, and the faults are all single-phase-to-earth ones. The fault causes include lightning, wildfire, crane collision and tree contact.
3.1 Hilbert marginal spectral feature

The Hilbert marginal spectrum $H(f)$ is obtained by integrating the Hilbert time-frequency spectrum $H(f, t)$ over the time domain. It reflects the statistical amplitude distribution of the data in the frequency domain, compared with the Fourier spectrum [4,5]. Typical marginal spectrums of the above four faults are shown in Figure 1.
The fault current frequency of crane collision is mainly concentrated at 50Hz, which means it is a metallic ground fault. The lightning fault is also approximate a metallic ground fault, but its DC content is higher, almost no high frequency harmonics. The fault resistance of wildfire and tree-caused fault are nonlinear. Therefore, their fault currents contain high frequency harmonics. Besides, the DC content is less.

3.2 Hilbert time-frequency spectral feature
The Hilbert time-frequency spectrum clearly indicates the relationship between signal frequency and time. The abscissa reflects the distribution of the signal in the time domain and the ordinate reflects its distribution in the frequency domain[7,8]. Typical time-frequency spectrums of four kinds of faults are shown in figure 2.
The time-frequency spectrums of crane collision and lighting fault are pretty similar. The signal energy during the fault is mainly concentrated at 50Hz. The wildfire fault has high frequency components during the whole fault, which may be related to the continuous combustion of the flame. The high-frequency components of the tree-caused fault are mainly distributed at the beginning and the end of the spectrum.
3.3 Spectral feature quantization
Take the ratio of DC component to the fundamental component as DC content, and the third harmonic to the fundamental component as harmonic content. The more dispersed the time-frequency spectrum, the larger the information entropy is. Therefore, we can use information entropy of image to describe the feature of time-frequency spectrums. Quantitative representation of features is shown in Table 1.

| Fault cause     | DC content | Harmonic content | Entropy |
|-----------------|------------|------------------|---------|
| Crane collision | <10%       | <5%              | 1.75-1.85 |
| Lighting        | 20%-40%    | <10%             | 1.8-1.9 |
| Wildfire        | <10%       | 10%-25%          | > 2     |
| Tree contact    | 5%-15%     | 5%-20%           | 1.8-1.9 |

4. PNN and classification results

4.1 PNN network structure
PNN is a branch of radial basis network. In essence, it is a supervised neural network classifier based on Bayesian minimum risk criterion. It has the advantages of fast learning process, classification and good fault tolerance.

PNN is generally composed of a four-layer structure: an input layer, a pattern layer, a summation layer, and an output layer. The input layer is responsible for passing the input feature vector to the network. The pattern layer calculates the degree of matching between the input feature vector and each mode in the training sample. Then the Euclidean distance is sent to the Gaussian function to obtain the pattern layer output. The summation layer is responsible for connecting the pattern layer units of each class. The output layer outputs the category with the highest score in the summation layer [13-15]. In this paper, three kinds of features are used to form feature vectors, and there are four fault categories. The corresponding PNN network structure is shown in Figure 3.

4.2 Algorithm process of PNN
①Normalize the training samples. Suppose there are m training samples, the training sample matrix is

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

The normalized coefficient matrix is

$$B = \frac{1}{\sqrt{\sum_{k=1}^{n} x_{1k}^2}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} \sum_{k=1}^{n} x_{1k}^2 \\ \sum_{k=1}^{n} x_{2k}^2 \\ \vdots \\ \sum_{k=1}^{n} x_{mk}^2 \end{bmatrix}$$

The normalized training sample is
\[ C_{m \times n} = B_{m \times 1} [1 \ 1 \ldots 1]_{1 \times n} \cdot X_{m \times n} \]  \hspace{1cm} (2)

② The Euclidean distance between the training sample matrix and the corresponding elements of the test sample matrix is

\[ E = \begin{bmatrix}
\sum_{k=1}^{n} |d_{1k} - c_{1k}|^2 \\
\sum_{k=1}^{n} |d_{1k} - c_{2k}|^2 \\
\vdots \\
\sum_{k=1}^{n} |d_{1k} - c_{mk}|^2 \\
\sum_{k=1}^{n} |d_{2k} - c_{1k}|^2 \\
\sum_{k=1}^{n} |d_{2k} - c_{2k}|^2 \\
\vdots \\
\sum_{k=1}^{n} |d_{2k} - c_{mk}|^2 \\
\vdots \\
\sum_{k=1}^{n} |d_{pk} - c_{1k}|^2 \\
\sum_{k=1}^{n} |d_{pk} - c_{2k}|^2 \\
\vdots \\
\sum_{k=1}^{n} |d_{pk} - c_{mk}|^2
\end{bmatrix} \hspace{1cm} (3)

\[ c_{ij} \] is the normalized training sample matrix element, and \( d_{ij} \) is the normalized element of the sample to be tested.

③ Activate the Gaussian function neuron to get the initial probability matrix (\( \sigma = 0.1 \)) as

\[ P = \begin{bmatrix}
e^{-\frac{E_{11}}{2\alpha^2}} & e^{-\frac{E_{12}}{2\alpha^2}} & \ldots & e^{-\frac{E_{1m}}{2\alpha^2}} \\
e^{-\frac{E_{21}}{2\alpha^2}} & e^{-\frac{E_{22}}{2\alpha^2}} & \ldots & e^{-\frac{E_{2m}}{2\alpha^2}} \\
\vdots & \vdots & \ldots & \vdots \\
e^{-\frac{E_{pl}}{2\alpha^2}} & e^{-\frac{E_{p2}}{2\alpha^2}} & \ldots & e^{-\frac{E_{pm}}{2\alpha^2}}
\end{bmatrix} \hspace{1cm} (4)

④ Suppose there are \( m \) samples, a total of \( c \) classes, the number of samples is the same as \( k \), and the initial probability that each sample of the summation layer belongs to each class is

\[ S = \begin{bmatrix}
\sum_{l=1}^{k} P_{1l} & \sum_{l=k+1}^{2k} P_{1l} & \ldots & \sum_{l=m-k+1}^{m} P_{1l} \\
\sum_{l=1}^{k} P_{2l} & \sum_{l=k+1}^{2k} P_{2l} & \ldots & \sum_{l=m-k+1}^{m} P_{2l} \\
\vdots & \vdots & \ldots & \vdots \\
\sum_{l=1}^{k} P_{pl} & \sum_{l=k+1}^{2k} P_{pl} & \ldots & \sum_{l=m-k+1}^{m} P_{pl}
\end{bmatrix} \hspace{1cm} (5)

⑤ Calculate the probability of fault type

\[ \text{prob}_{ij} = S_{ij} / \sum_{l=1}^{c} S_{il} \hspace{1cm} (6) \]

4.3 Identification results of fault cause

Trained by 76 samples and tested by 43 samples, the classification results by PNN are shown in Table 2. The recognition rates of the four faults are all above 80%, and the overall recognition rate is 88.4%.

| Fault cause       | Samples | Identified | Rate% |
|-------------------|---------|------------|-------|
| Crane collision   | 11      | 9          | 81.8  |
| Lighting          | 14      | 12         | 85.7  |
|                  |   |   |     |
|------------------|---|---|-----|
| Wildfire         | 9 | 9 | 100 |
| Tree contact     | 9 | 8 | 88.9|
| Sum              | 43| 38| 88.4|

5. Conclusion
This paper describes the principle of HHT and analyzes the zero-sequence current, which will result in the Hilbert marginal spectrum and Hilbert time-frequency spectrum used as features. On this basis, PNN for classification and identification has achieved the expected goals. In the follow-up work, it is necessary to dig deeper into other fault features, optimize the representation of features, and identify faults types not covered in this paper effectively.

References
[1] J. Wang, Q. Yang, L. Chen, W. Sima, High Voltage Engineering, 38, 2068 (2012)
[2] Y. Hu, K. Liu, T. Wu, Y. Liu, Z. Su, High Voltage Engineering, 40, 3491 (2014)
[3] T. Wu, Y. Hu, J. Ruan, High Voltage Engineering, 37, 1115 (2011)
[4] X. Peng, X. Li, S. Yao, G. Qian, Southern Power System Technology, 6, 43 (2012)
[5] H. Li, Y. Zhang, J. Yu, Hydropower Automation And Dam Monitoring, 38, 67 (2014)
[6] L. Lin, H. Yan, X. Yue, F. Hu, Southern Power System Technology, 8, 47 (2008)
[7] L. Yang, Y. Zhu, Power System Protection and Control, 43, 62 (2015)
[8] W. Zhang, C. Wang, Automation of Electric Power Systems, 31, 34 (2007)