Research on Open Circuit Fault Diagnosis of Inverter Circuit Switching tube Based on Machine Learning Algorithm

Huaqing Xu a, Yanqing Peng b, * and Lumei Su c

School of electrical engineering and automation, Xiamen University of Technology, Xiamen Fujian 361024, China

a huaqingxu89@gmail.com, *, b Corresponding author e-mail pyqxm@163.com, c sulumei@xmut.edu.cn

Abstract. The inverter circuit is widely used, and the switching tube has the highest incidence of open circuit failure. In order to improve the recognition rate of the open circuit fault of the inverter circuit, this paper proposes a method to extract the output voltage, output current and input current time domain features and then use the Random Forest and K-Nearest Neighbor to identify the fault. In this paper, through the simulation of single-phase full-bridge inverter circuit, the open-circuit fault test of the switch tube is simulated, and the output voltage, output current and DC-side input current of each switch tube open-circuit fault are obtained, and then the time domain feature extraction is performed. The Random Forest and K-Nearest Neighbor are used for diagnostic comparison. The results show that the single-tube fault recognition rate can reach more than 96% and the highest fault recognition rate can reach 99.77%.

Keywords: Simulation, Inverter Circuit, Fault Diagnosis, K-Nearest Neighbor, Random Forest.

1. Introduction

Inverter circuit is a kind of circuit structure that can convert DC power into AC power. It is used in many important power electronic equipment, such as uninterruptible power supply, wind power generation, electric vehicles and so on. The core technology is Pulse-Width Modulation (PWM) technology. The most common one is Sine Pulse-Width Modulation (SPWM) technology. The sine wave is used as the modulation wave and controlled according to the sine time interval. The conduction state of the switch tube enables the inverter circuit to output alternating current. As the core component of the inverter circuit, the switch tube has a large open circuit failure rate. Therefore, the fault diagnosis and positioning of the open circuit failure of the inverter circuit has practical significance and practical value.

At home and abroad, the research on the fault diagnosis of the open circuit of the inverter circuit is as follows: [1] using the fast Fourier analysis algorithm to analyze and diagnose the current harmonic at the output of the inverter circuit under linear and nonlinear load conditions; In [2], the wavelet analysis and support vector machine are used to diagnose the fault of the three-phase PWM inverter circuit. The author first uses the discrete wavelet transform to detect the current discontinuity caused by the fault, so as to extract the fault features and then train with these features. Support vector machines and classify faults, and their effectiveness is verified; In [3], a three-dimensional feature extraction method is
proposed, and the artificial neural network is used to diagnose the fault of the three-phase inverter circuit. The accuracy and reliability of the system are verified by simulation experiments; The literature [4] is based on hybrid support vector machine and Wavelet analysis proposes a fault classification method for inverter circuits. Firstly, the discrete voltage wavelet transform is used to decompose the output voltage waveform of the inverter circuit to obtain the wavelet coefficient matrix. The wavelet coefficient matrix can be used to obtain the singular value vector as the characteristic timing period waveform, and then use the hybrid model of Huffman tree and SVM to classify faults. The verified correct rate is 99.6%, which is 3.65% higher than ordinary SVM; Literature [5] proposes a fault characteristic parameter preprocessing method suitable for three-phase inverter circuit, and processes the characteristic signals collected by the existing sensors of the inverter circuit, establishes the output line voltage envelope function and the diagnostic function, and utilizes the line voltage envelope. And fault parameter characteristics, accurately locate the open circuit fault of the switch tube, and provide reliable fault information for fault-tolerant control and quick maintenance of the inverter system; In [6], a fast diagnosis and localization method for single-switch open-circuit fault of three-phase inverter circuit based on current vector characteristic analysis is proposed, The fault vector is diagnosed rapidly by the change of current vector instantaneous frequency, combined with the current vector instantaneous angle information, Fault phase localization, combined with a current average based diagnostic variable for fault switch tube positioning, through simulation and build hardware experimental platform to prove that the proposed fault diagnosis and positioning method can effectively locate the fault phase and fault switch tube; In [7], an active power filter IGBT fault feature extraction method based on multi-feature fusion is proposed. This method collects the clamp diode arm voltage in the three-level APF main circuit as the test signal, and performs wavelet decomposition and extraction. The energy coefficient, power spectral entropy and singular spectral entropy of each frequency band are used to form a multi-feature parameter matrix, and then feature reduction is performed to form a eigenvector matrix. Experimental results show that the multi-feature fusion extraction method overcomes the one-sidedness of single features. The feature extraction method of [1-6] is difficult, and the extraction features are few, which is easy to reduce the recognition accuracy due to insufficient data information. The literature [7] uses frequency domain multi-feature extraction, but analysis and conversion are a bit cumbersome between time domain and frequency domain.

The feature extraction methods commonly used for continuous time series are time domain feature extraction and frequency domain feature extraction [8]. The time domain features are divided into dimensioned time domain features and dimensionless time domain features. Among them, the common dimension time domain features are: variance, standard deviation, maximum value, minimum value, mean value, mean square value, peak value, mean square amplitude, average amplitude, square root amplitude, peak-to-peak value, etc. common dimensionless time domain features are: margin index, pulse index, peak index, waveform index, kurtosis index, skewness index and so on. The frequency domain features are mainly: harmonic factor, correlation factor, power spectrum center of gravity index, spectrum origin moment, power spectrum variance and so on.

In this paper, 12 time-domain features are selected to extract the output voltage, output current and DC-side input current. Finally, a data set containing 36 features is obtained. Random Forest and K-Nearest Neighbor are used for diagnosis. The result shows the fault recognition rate can reach more than 96% and the highest can reach 99.77%. The experimental process is shown in Figure 1.
2. Simulation model of inverter circuit

The experiment uses a single-phase full-bridge inverter circuit, as shown in Figure 2. Among them, S1, S2, S3 and S4 are insulated gate bipolar transistors (IGBT), which adopt sinusoidal pulse width modulation (SPWM) technology, and use narrow pulse with varying width as the driving signal. The theoretical basis is the area equivalent principle, that is, the shape is different. When the narrow pulse of equal area is added to the linear inertia link, the output effect is basically the same, which is a kind of inverter control technology which is currently cited more.

SPWM technology is divided into bipolar SPWM technology and unipolar SPWM technology. The output voltage of the inverter circuit under bipolar SPWM technology has positive and negative levels. The output voltage of the inverter circuit under unipolar SPWM technology has zero and positive (or negative) levels. This experiment uses bipolar SPWM technology.

This paper uses Simulink for simulation experiment. Bipolar SPWM signal generation diagram shown in Figure 3. Bipolar SPWM Single-phase full-bridge inverter circuit diagram shown in Figure 2.
Among them, DC voltage is 220V, output resistance is $1 \Omega$, output inductance is 2mH, internal resistance of IGBT is $1m \Omega$, snubber resistance of IGBT is 100000 $\Omega$, snubber capacitance of IGBT is inf F, modulation depth $m$ is 0.5, modulation wave frequency is 50Hz and carrier wave frequency is 1500Hz. Open circuit failure test on all four switching tubes, and then the output signal is superimposed with Gaussian white noise with a signal-to-noise ratio of 27, and the output signal waveform is as shown in Fig. 4.
3. **Time domain features extraction**

Three cycles of data were collected for the normal conditions, S1 open circuit fault, S2 open circuit fault, S3 open circuit fault, and S4 open circuit fault. Each category has 600 data points and a total of 3000 data points. The data was subjected to one-time five-point three smoothing operation and normalized, and 12 common time-domain features were selected to perform time-domain feature extraction on output voltage, output current and input current. The selected 12 kinds of time domain features are shown in Table 1.

| Variances | Maxima | Minimums | Means |
|-----------|--------|----------|-------|
| $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$ | $x_{\max} = \max_{i \in \{1,2,\ldots,N\}} \{x_i\}$ | $x_{\min} = \min_{i \in \{1,2,\ldots,N\}} \{x_i\}$ | $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ |
| P-P value | Standard deviation | Kurtosis | Skewness |
| $x_{p-p} = x_{\max} - x_{\min}$ | $\sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$ | $x_k = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{x_i - \bar{x}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}} \right)^4$ | $x_s = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{x_i - \bar{x}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}} \right)^3$ |
| Waveform | Peak | Pulse | Margin |
| $K = \frac{1}{N} \sum_{i=1}^{N} x_i^2 / \frac{1}{M} \sum_{m=1}^{M} x_m$ | $C = x_{\max} / \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}$ | $I = x_{\max} / \left( \frac{1}{M} \sum_{m=1}^{M} |x_m| \right)$ | $CL_i = x_{\max} \left( \frac{1}{M} \sum_{m=1}^{M} |x_m| \right)^2$ |

Set the sliding window width to the data points (200) included in one cycle of the output signal, the step size is 1 data point, and perform 12 kinds of time domain feature extractions for output voltage, output current and input current.
output current and input current in each sliding window. Finally, a data set with 36 features and 2801 samples is obtained. Correspondingly, the data set sample category is set to five categories, and the category label is named "class0", "class1", "class2", "class3" and "class4" according to "normal working condition", "S1 open circuit fault", "S2 open circuit fault", "S3 open circuit fault" and "S4 open circuit fault". The total number of samples is 2801. In order to facilitate the experimental operation, the last sample of the class4 is deleted. The distribution of the final samples class0 to class4 is as follows: 401, 600, 600, 600, 599, totaling 2800.

4. Fault diagnosis experiment

In this paper, the K-Nearest Neighbor and Random Forest Algorithm are used to diagnose the open circuit fault of the inverter circuit. Both the K-Nearest Neighbor Algorithm and the Random Forest Algorithm are supervised learning algorithms. The K-Nearest Neighbor Algorithm is the Nearest Neighbor Algorithm when K is taken as 1. The input sample is discriminated as the class of the nearest sample. When K is not equal to 1, the input is discriminated as the great majority sample category among the K samples; the Random Forest Algorithm [9] is an integrated learning algorithm, it is widely used in various fields, such as data mining [10] and computer vision [11-14], and so on. The base classifier is a Classification and Regression Tree (CART). Each training sample set of CART in the Random Forest uses sampling with replacement, and the feature set uses sampling without replacement. When the number of CART is enough, the Random Forest can show very good prediction performance.

The experiment uses a 10-fold Cross-Validation method to take the mean value of 100 repeated experiments. The open circuit fault recognition rate and overall classification correct rate of each switch are shown in Table 2.

| Fault Diagnosis Algorithm | S1 Fault Recognition Rate [%] | S2 Fault Recognition Rate [%] | S3 Fault Recognition Rate [%] | S4 Fault Recognition Rate [%] | Correct Classification Rate [%] |
|---------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|
| KNN(K=1)                  | 99.54                         | 96.58                         | 99.39                         | 96.64                         | 97.58                          |
| KNN(K=5)                  | 99.50                         | 96.79                         | 99.35                         | 96.85                         | 98.38                          |
| Random Forest (ntree=1)   | 98.58                         | 98.52                         | 98.73                         | 98.78                         | 99.10                          |
| Random Forest (ntree=100) | 99.55                         | 99.55                         | 99.50                         | 99.77                         | 99.77                          |

It can be seen from Table 2 that the lowest fault recognition rate can reach more than 96%, and the highest can reach 99.77%. In general, the Random Forest performance is optimal when the CART number is 100.

5. Conclusion

In this paper, Simulink is used to simulate the open circuit fault of the inverter circuit, and the output voltage, output current and input current are obtained when the single-tube open-circuit fault is simulated, and Gaussian white noise is added appropriately. Then, after data smoothing and normalization processing, 12 common time domain features are extracted for output voltage, output current and input current. The sliding window extraction method is used to obtain 36 features and 2801 samples data sets. Finally, the K-Nearest Neighbor and Random Forest algorithm are used for fault diagnosis. The results show that the Random Forest has the best performance when the number of CART is 100, and the highest fault recognition rate can reach 99.77%.

Acknowledgements

This research was financially supported by the Fujian Province Science Foundation(2017J01510).
References

[1] Shim, Won-Sul, et al. "A real-time fault diagnosis system for UPS based on FFT frequency analysis." Applied Power Electronics Conference and Exposition (APEC), 2010 Twenty-Fifth Annual IEEE. IEEE, 2010.

[2] Kim, Dong-Eok, and Dong-Choon Lee. "Fault diagnosis of three-phase PWM inverters using wavelet and SVM." Journal of Power Electronics 9.3 (2009): 377-385.

[3] Talha, Muhammad, Furqan Asghar, and Sung Ho Kim. "A Novel Three-Phase Inverter Fault Diagnosis System Using Three-dimensional Feature Extraction and Neural Network." Arabian Journal for Science and Engineering (2018): 1-14.

[4] Hu, Zhi-kun, et al. "Fault classification method for inverter based on hybrid support vector machines and wavelet analysis." International Journal of Control, Automation and Systems 9.4 (2011): 797.

[5] Chen, Gaohua, et al. "A Non-Intrusive Diagnostic Technique for Open-Circuited Faults of Inverters." Proceedings of the Csee37.13(2017):3854-3862.

[6] Chen, Yong, Z. Liu, and Z. Chen. "Fast Diagnosis and Location Method for Open-Circuit Fault in Inverter Based on Current Vector Character Analysis." Transactions of China Electrotechnical Society (2018).

[7] Sun, Shuguang, et al. "Feature extraction of IGBT open circuit fault in active power filter based on multi-feature fusion." Chinese Journal of Scientific Instrument (2017).

[8] Samimy, Bahman, and G. Rizzoni. "Mechanical signature analysis using time-frequency signal processing: application to internal combustion engine knock detection." Proceedings of the IEEE 84.9(1996):1330-1343.

[9] Breiman, Leo. "Random Forests." Machine Learning 45.1(2001):5-32.

[10] Verikas, A, A. Gelzinis, and M. Bacauskiene. "Mining data with random forests: A survey and results of new tests." Pattern Recognition44.2(2011):330-349.

[11] Shotton, Jamie, et al. "Real-time human pose recognition in parts from single depth images." Communications of the ACM 56.1(2013):116-124.

[12] Gall, J., and V. Lempitsky. "Class-Specific Hough Forests for Object Detection." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on IEEE, 2013:1022-1029.

[13] Ren, Shaoqing, et al. "Face Alignment at 3000 FPS via Regressing Local Binary Features." IEEE Conference on Computer Vision and Pattern Recognition IEEE Computer Society, 2014:1685-1692.

[14] Ren, Shaoqing, et al. "Global refinement of random forest." IEEE Conference on Computer Vision and Pattern Recognition IEEE Computer Society, 2015:723-730.