Study of the Unsupervised Extraction Method of Transformer Vibration Features

Wei Xu[1], Chi Kang[1], Xiangjing Du[1], Bin Zhong[1], Naihui Wang[1], Zhong Jin[1] and Yang Jing[2,*]
[1]Leshan Power Supply Company, State Grid Sichuan Electric Power Corporation, Leshan, 614000, China
[2]School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China
[*]Corresponding author

***Abstract*—A mutual information-based unsupervised feature extraction method of transformer surface vibration is proposed in the paper. Wavelet packet analysis is used to extract surface vibration signal frequency band-energy features of the transform in operation and complete its signal representation. Relevance and redundancy of mutual information measurement are considered to judge importance degrees of features, and feature importance ranking and selection are completed based on unsupervised minimum redundancy and maximum relevance criteria. The analysis of measured signals at different measuring points of the transformer indicates that this method can accurately select important features of transformer vibration signals while effectively reducing data dimensionalities.**

***Keywords*—transformer; vibration signal; feature selection; wavelet packet; mutual information**

I. INTRODUCTION

Transformer is an important equipment in the electric power system. Theoretical analysis and practical experience indicate that its operating state and fault diagnosis can be analyzed through its vibration signals [1].

Transformer surface vibration mainly derives from iron core vibration influenced by internal magnetostriction and winding vibration caused by electromagnetic force, and cooling systems (fans, oil pumps, etc.) and voltage regulating devices will also have a certain influence. Literature [2] used the transformer vibration testing system to study iron core vibration under idle condition and winding vibration under load conditions Literature [3] analyzed measured vibration signals of the transformer under a sudden short circuit based on complex wavelet transform and captured spectral features of transformer vibration signals under a sudden short circuit. Literature [4] put forward an improved empirical mode decomposition (EMD) algorithm to identify transformer winding mode parameters for the sake of vibration analysis of winding deformation.

Transformer surface vibration is influenced by multiple factors. Transformer surface vibration signals in actual operation are obviously different from those ac-quired through theoretical analysis and under laboratory test conditions. Literature [5] designed and constructed a portable vibration acquisition system to acquire trans-former surface vibration signals in operation, gave a GRNN-based calculation method of fundamental frequency amplitude and calculated fundamental frequency amplitude of surface vibration based on historical da-ta and combining operating conditions. Literature [6] studied and analyzed transformer surface vibration signal features under different working conditions and proposed a sensitivity index to quantitatively describe abilities of different frequency points of vibration signals to reflect winding vibration change. On the whole, there are many complicated influence factors of transformer surface vibration in operation, theoretical modeling for accurate analysis is difficult and even cannot be realized, and there are no reliable abnormality detection and fault diagnosis algorithms of transformer vibration.

Online monitored transformer vibration data presents linear growth with time, and effective feature ex-traction, highly efficient storage and accurate analysis constitute a technical difficulty. Meanwhile, transformer surface vibration data samples of different models under different working conditions are not complete, especially vibration data samples under transformer operation ab-normality and under various fault states are scarce, which brings about difficulty and challenge to feature extraction of transformer surface vibration.

II. THEORETICAL BASIS

A. Wavelet Packet Analysis-based Frequency Band-energy Representation

Different wavelet basis functions selected, waveform differences after decomposition are greatly different, so are signal analysis results; the greater the decomposition layers, the higher the frequency band division accuracy, and it's conductive to accurate extraction of signal energies at frequency bands, but the computational complexity increases, which goes against follow-up analysis and processing.

B. Mutual Information-based Unsupervised Feature Selection

Monitored transformer data size is enormous and contains a large quantity of abundant and even irrelevant data with few abnormalities or fault data, which makes it difficult for feature extraction of transformer vibration signals. With a reference to the method in Literature [7], relevance and redundancy between features are firstly analyzed using mutual information and then feature importance ranking is implemented based on maximum relevance and minimum redundancy criteria to complete feature selection.

1) Mutual Information Principle: In the information entropy theory of Shannon, entropy is used to measure uncertainty between variable information. H(X) is usually used to describe entropy of discrete variables X= $\{x_1, x_2, \dots x_n\}$, where x_i is possible value of variable X; $p(x_i)$ is probability density function. X and Y are set as two discrete random variables and p(X) is probability for variable X to take different values, and then uncertainty of variable X value can be expressed by information entropy:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) \tag{1}$$

On the precondition of different known possible values of a variable *Y*, uncertainty of variable *X* value is expressed by conditional entropy:

$$H(X \mid Y) = -\sum_{j=1}^{m}\{p(y_j)\times \sum_{i=1}^{n} p(x_i \mid y_j)\log p(x_i \mid y_j)\} \tag{2}$$

Where x_i and y_j are different possible values of *X* and *Y* respectively. When two variables are mutually independent, then conditional entropy is equal to information entropy, and as a general rule, information entropy is greater than conditional entropy.

Mutual information between two random variables *X* and *Y* is defined as:

$$I(X;Y) = H(X) - H(X \mid Y) = I(Y;X) \tag{3}$$

Mutual information can express information quantity contained in both variables *X* and *Y*, it can be regarded as reduction of uncertainty of a variable *X* when concrete information of another variable *Y* is known, its value is symmetrically nonnegative, and the greater the value, the higher the relevance degree, and complete irrelevance exists when its value is 0.

2) Feature Sorting and Selection: In the feature selection process, it's necessary to exclude redundant features whole selecting related features. Definition and calculation method of "relevance" and "redundancy" of features to be selected and the whole feature set are given as follows.

Definition 1 Relevance: relevance of feature f_i refers to average mutual information of it with the whole feature set.

$$\begin{aligned} \mathrm{Re}l(f_i) &= \frac{1}{n}\sum_{j=1}^{n} I(f_i; f_j) \\ &= \frac{1}{n}\left(H(f_i) + \sum_{1\le j\le n, j\ne i} I(f_i; f_j)\right) \end{aligned} \tag{4}$$

Where $H(f_i)$ is information quantity contained in feature f_i. The greater the $H(f_i)$ value, the larger the information quantity contained in feature f_i. $\sum_{1\le j\le n, j\ne i} I(f_i; f_j)$ is information quantity jointly contained in feature f_i and other features, and the greater the value, the fewer the "new" knowledge which can be provided by other features to the set.

Assuming a to-be-selected feature f_i in the to-be-selected feature set *U*, conditional information entropy of any selected feature g_t in the selected feature set S_{m-1} relative to to-be-selected feature f_i is $H(g_t|f_i)$. If to-be-selected feature f_i is added to selected feature set S_{m-1}, then information quantity provided by selected feature g_t is reduced due to addition of to-be-selected feature f_i, and relevance of selected feature g_t itself is reduced, so "conditional relevance" is defined as follow.

Definition 2 Conditional relevance: conditional relevance of selected feature g_t relative to to-be-selected feature f_i is defined as:

$$\mathrm{Re}l(g_t \mid f_i) = \frac{H(g_t \mid f_i)}{H(g_t)}\mathrm{Re}l(g_t) \tag{5}$$

According to the above equation, conditional relevance is not greater than relevance. Their difference can be defined as redundancy naturally only when two features are mutually independent and equal.

Definition 3 Redundancy: redundancy of to-be-selected feature f_i to selected feature g_t is defined as difference value between relevance $Rel(g_t)$ of selected feature g_t and its conditional relevance $Rel(g_t|f_i)$.

$$\mathrm{Re}d(f_i; g_t) = \mathrm{Re}l(g_t) - \mathrm{Re}l(g_t \mid f_i) \tag{6}$$

Based on the above definition, unsupervised maximal relevance and minimal redundancy (UmRMR) is obtained [8], where the core idea is to seek for the feature having the maximum "relevance" with the whole set and minimum "redundancy" to the selected feature set, and its expression is as follow:

$$UmRMR(f_i) = \mathrm{Re}l(f_i) - \max_{g_t\in S_{m-1}}\left\{\mathrm{Re}d(f_i; g_t)\right\} \tag{7}$$

The following is set

$$l_m = \arg\max_{1\le i\le n}\left\{UmRMR(f_i) \mid f_i \in U\right\} \quad (8)$$

The m(th) feature can be selected as $g_m=f_{lm}$, as this feature reduces uncertainty of other features to the greatest extent and brings the minimum redundant information.

III. FEATURE EXTRACTION OF TRANSFORMER SURFACE VIBRATION SIGNALS

Surface vibration data of a normally operating transformer in a transformer substation are acquired, and transformer model is SFZ9-50000/110. 6 acceleration sensors are adsorbed on positions 1/3 away from the transformer bottom at high-voltage-side phase A, B and C and low-voltage-side phase a, b and c respectively, and measuring points are numbered as 1-6 successively. Sampling frequency is 10kHz, recording duration in each sampling process is 1s and sampling interval is 5min.

Transformer surface vibration signals acquired on the field are organized, and frequency spectral records at 6 measuring points at the same moment are shown in FIGURE I.

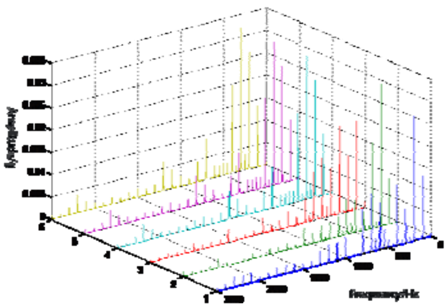


FIGURE I. SPECTRA CHART OF TRANSFORMER SURFACE VIBRATION SIGNAL AT DIFFERENT MEASUREMENT POSITIONS

A. Frequency Band-energy Representation

Daubechies1 wavelet is used to complete frequency band-energy representation of transformer vibration signals. In consideration of transformer vibration signal features and follow-up processing demand, the wavelet packet is decomposed into 6 layers, bandwidth of each section is 78.125Hz and frequency band division at the 6^{th} layer is shown in TABLE I.

TABLE I. WAVELET PACKET DECOMPOSITION FREQUENCY RANGE OF EACH FREQUENCY BAND (UNIT: HZ)

Frequency band	1	2	⋯	63	64
Frequency Range	0~ 78.125	78.125~ 156.25	⋯	4843.75~ 4921.875	4921.87 5~5000

In view that transformer surface vibration energy is mainly distributed below 1,500Hz (FIGURE I), and processing steps of vibration signal frequency band-energy representation are as follows: wavelet packet decomposes the transformer surface vibration signal, wavelet basis function is Daubechies1, and number of decomposition layers is 6; signal energies of frequency bands 1-20 (TABLE I) are calculated and recorded as $E_1,E_2,\ldots,E_{20}$ in succession.

B. Data Preprocessing

Normalization processing of energy value at each frequency band is carried out. Transformer vibration signal frequency band-energy eigenvector after normalization processing is recorded as $X=[x_1,x_2,\ldots,x_{20}]$, where:

$$x_i = E_i \Big/ \sum_{k=1}^{20} E_k \quad (9)$$

Frequency band-energy data obtained through normalization are continuous variables, and it's difficult to calculate their probability distribution and mutual information. Therefore, it's also necessary to conduct discretization. Unsupervised equal-width discrete method is used, number of distribution boxes is set as 10, and transformer vibration signal eigenvector consisting of 20 features is obtained after processing and recorded as $V=[f_1,f_2,\ldots,f_{20}]$.

C. Mutual Information-based Unsupervised Feature Sorting

The method specified in section 1.2 is used to analyze importance degrees of dimensional features of vibration signal eigenvector V and conduct sorting. $I(f_i;f_j)$ is defined as mutual information value between two features in V; $H(f_i;f_j)$ is conditional entropy of feature f_i; $H(f_i)$ is f_i feature entropy; S is the selected feature set after sorting; U is to-be-selected feature set, and the sorting algorithm is described as below:

(1) Initialization: $S=\varnothing$, $U=D(f_1,f_2,\cdots,f_n)$;

(2) Mutual information calculation: for any $f_i, f_t \in U$, $H(f_i), H(f_i \mid f_j), I(f_i;f_j)$ are calculated;

(3) The first feature is selected:

The selected first feature should be the one which can provide the maximum information quantity and reduce uncertainty of other features to the greatest extent. Selection of the first feature is decided by the following equation:

$$score(f_i) = \frac{1}{n}\sum_{j=1}^{n} I(f_i;f_j) \quad (10)$$

$$l_1 = \arg\max_{1\le i\le n}\left\{score(f_i)\right\} \quad (11)$$

$s_1 = f_{l_1}$, $S=\{f_{l_1}\}$ and $U=U \setminus f_{l_1}$ are set.

(4) Selection of other features: feature f_{l_m} which satisfy formulas $UmRMR(f_i) = \mathrm{Re}\, l(f_i) - \max_{g_t \in S_{m-1}} \{\mathrm{Re}\, d(f_i; g_t)\}$ and $l_m = \arg\max_{1 \le i \le n} \{UmRMR(f_i) \mid f_i \in U\}$ is selected from U, and $S \bigcup \{f_{l_m}\} \to S$ and $U \setminus f_{l_m} \to U$ are set.

(5) Step (4) is repeated until $U = \varnothing$, and S is output.

Finally output S of the algorithm are orderly feature sequences sorted according to feature importance.

The transformer vibration signal is processed according to the abovementioned method so as to obtain orderly feature sequences sorted according to feature importance. TABLE II lists data feature sorting results at high-voltage side phase A, B and C and low-voltage side phase a (recorded as HA, HB, HC and LA successively).

TABLE II. TRANSFORMER SURFACE VIBRATION SIGNAL UNSUPERVISED FEATURE SELECTION RESULT

NO.	Data Set	Features ranking by UFS
1	HA	{08,18,02,07,05,11,19,04,09,20,13,10,06,17,16,01,03,14,15,12}
2	HB	{08,18,07,09,01,14,19,04,02,17,15,13,05,03,20,16,12,06,10,11}
3	HC	{03,02,04,05,08,20,13,14,17,10,06,16,09,19,18,07,01,11,15,12}
4	LA	{06,02,05,20,07,10,11,19,16,04,17,08,13,18,01,14,12,15,09,03}

IV. CLASSIFICATION TEST

In order to analyze whether the method proposed in this paper can identify important transformer surface vibration features, relevance or dependence between generated orderly feature sequences and transformer measuring points (namely potential categories) and whether it can reduce data dimensionalities and improve the follow-up algorithm performance, K-nearest neighbor classification algorithm is used to conduct the classification test, and classification accuracy is taken as the index evaluating performance of the feature selection method.

D. Experimental Data

Measuring points are taken as the category division basis to process transformer surface vibration data for continuous 20h, 6 categories of sequential feature datasets (238 records in each category) of the transformer vibration signal (totally 1,428 records) are obtained after data preprocessing, and recorded as Dataset A.

With a reference to feature sorting results of different dimensionalities at high-voltage side phase C in TABLE III, data in Dataset A are re-sorted to obtain orderly feature sequences sorted according to importance degree from high to low, and the obtained data are recorded as Dataset B.

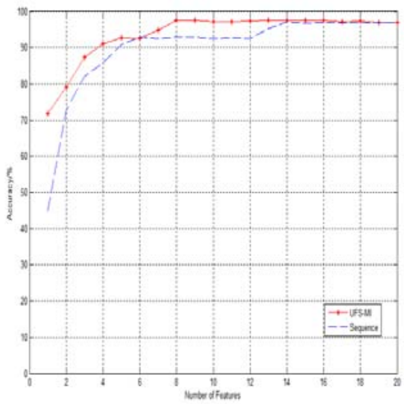
E. Classification Test

Feature selection can be regarded as seeking for a feature subset containing original features and integrating most or all information. This feature subset can reduce uncertainty of other unselected features to the greatest extent.

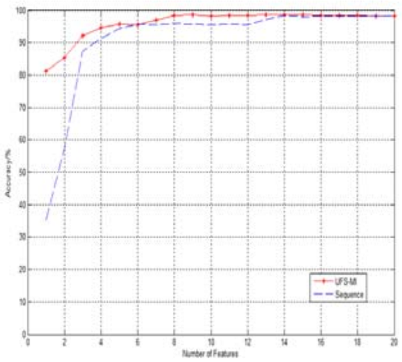
According to feature sequence in 2 experimental datasets, features present orderly progressive increase to constitute a feature subset, data characterized by this feature subset is taken as input of K-nearest classier, K value is defined as 1, 3 and 5 respectively for classification, and average value of operating results of 10-fold cross validation for 10 times is selected as the final classification accuracy. In consideration that similarity of surface vibration at the same transformer side (high-voltage side or low-voltage side) is high while difference between different sides is large, the experiment is carried out in 2 parts.

Vibration data at three measuring points—high-voltage side phase A, B and C—with high similarity degree in experimental data are selected in the first part to conduct K-nearest classification. FIGURE II gives change tendencies of classification accuracy of the classier when different numbers of features are selected to characterize data, where subgraphs (a), (b) and (c) show classification results when K value is taken as 1, 3 and 5 respectively. Data classification result in Dataset A is marked as Sequence, and that in Dataset B is marked as UFS-MI.

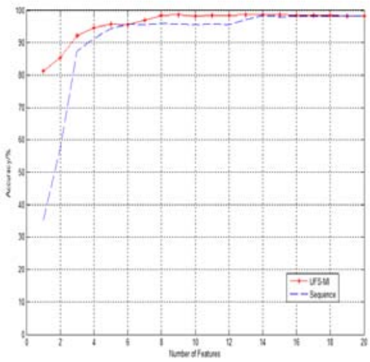
The similar method is used in the second part to conduct classification test in the whole experimental dataset. FIGURE III gives change tendencies of nearest classification accuracy of data at six measuring points of the test transformer under different K values, and the optimal classification accuracy is recorded in TABLE III.



(a)K=1

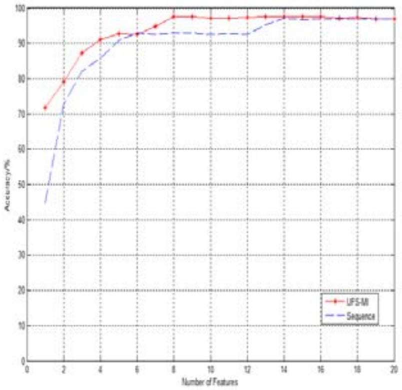


(b)K=3

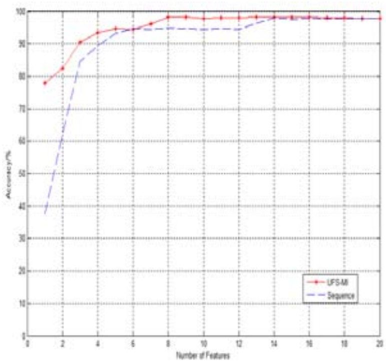


(c)K=5

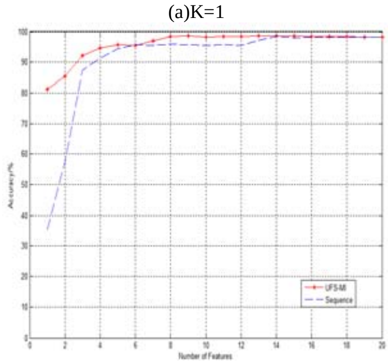
FIGURE II. COMPARISON OF CLASSIFICATION PERFORMANCE OF DIFFERENT VIBRATION FEATURES DATA



(a)K=1



(b)K=3



(c)K=5

FIGURE III. FIGURE III. COMPARISON OF CLASSIFICATION PERFORMANCE OF DIFFERENT VIBRATION FEATURES DATA

TABLE III. TWO METHODS OF DIFFERENT SIZE FEATURE SUBSET CLASSIFICATION ACCURACY

No. Features	Accuracy (%)		No. Features	Accuracy (%)	
	UFS-MI	*Sequence*		*UFS-MI*	*Sequence*
1	84.42	46.16	11	98.62	96.38
2	87.90	64.57	12	98.70	96.30
3	93.48	89.42	13	**98.84**	97.61
4	95.58	92.83	14	**98.84**	**98.62**
5	96.38	95.43	15	**98.84**	98.33
6	96.30	96.45	16	98.77	98.48
7	97.39	96.30	17	98.62	98.48
8	98.77	96.52	18	98.70	98.48
9	**98.84**	96.45	19	98.48	98.48
10	98.55	96.23	20	98.48	98.48

Notes: No. Features expresses number of features in the feature subset.

The experiment indicates that as the number of features increases, classification accuracy is improved and then tends to be a relatively stable value or to decline somehow, certifying that feature information initially input in the classifier is insufficient, and classification performance declines due to addition of redundant or irrelevant data.

Based on the feature sorting method, classification accuracy according to the first feature reaches above 70% and presents rapid progressive increase as the number of features increases, and the highest classification accuracy is reached when number of features is 8 or 9; classification accuracy of the first feature of each sequential feature data doesn’t reach 50%, and the highest classification accuracy is reached only when the number of features increases to 14. It’s proved that the method proposed in this paper can accurately recognize important features of transformer surface vibration, and feature dimensionalities reflecting important signal features are placed in front row of the feature sequence.

The highest classification accuracies on experimental datasets are 98.84% and 98.62% respectively. Under the same quantity of features, classification accuracies of both sorted feature data are higher than or equal to classification effect of sequential feature data.

V. CONCLUSIONS

An unsupervised feature extraction method of transformation surface vibration is given in this paper, and conclusions are drawn as follows: For the demand for feature extraction of the transformer surface vibration signal, wavelet packet analysis-based hierarchical division basis of the transformer surface vibration signal and its frequency band-energy feature representation method is given. Based on relevance and redundancy between mutual information-based measured features and combining unsupervised maximum relevance and minimum redundancy criteria, a feature extraction method of the transformer surface vibration signal is given. Classification test of measured transformer vibration data indicates that the method proposed in this paper can acquire classification performance superior to that on a complete feature dataset while reducing half of features.

REFERENCES

[1] C. Bartoletti, M. Desiderio, D. Di Carlo, G. Fazio, F. Muzi, G. Sacerdoti and et al., "Vibro-acoustic techniques to diagnose power transformers," IEEE Transactions on Power Delivery, vol. 19, no. 1, pp. 221-229, January 2004..

[2] S. Ji, J. Cheng, and Y. Li, "Research on Vibration Characteristics of Windings and Core of Oil-Filled Transformer" Jou Rnal Of Xi'An Jiao Tong University, Vol. 39 No.6,pp. 616-619+658,Jun.2005

[3] K. Zhang, F. H. Wang, T.M. Ming, and Z.J. Jin, "Detection of transformer winding deformation under sudden short-circuit impact based on complex wavelet algorithm," Transactions of China Electrotechnical Society, vol.29, no.8, pp. 327-332.

[4] C. Geng, F. Wang, J. Zhang, and Z. Jin, "Modal parameters identification of power transformer winding based on improved Empirical Mode Decomposition method," Electric Power Systems Research, vol.108, pp. 331-339.2014

[5] Z. Li, W. Zhang, N. Sun, and T. Song, "Calculation of Vibration Fundamental Frequency Amplitude of Transformer SurfaceBased on Generalized Regression Neural Network," High Voltage Engineering, Vol.43, No.7:pp. 2287-2293, July 31, 2017

[6] Z. Li, T. Song, T. Guo, and W. Zhang, "Characteristics analysis on three-dimensional surface vibration signals of transformer under different load current," Power System Protection and Control, Vol.45 No.12, pp. 29-34, Jun. 16, 2017

[7] S. Wu, W. Huang, F. Kong, Q. Wu, and P. Zhu, "Vibration features extraction of power transformer using an time-scale-frequency analysis method based on WPT and HHT," in 2009 IEEE 6th International Power Electronics and Motion Control Conference, IEEE, pp. 2577-2581. ,2009

[8] J. Xu, Y. Zhou, L. Chen, and B. Xu, "An unsupervised feature selection approach based on mutual information," Journal of Computer Research and Development, vol.49, no.2, pp. 372-382.2012