

## Fault Diagnosis of Power Quality and Disturbance Classification based on KPCA - FDA Method

---

### Liudan Xie

*Zhejiang Truly Electric Co., Ltd.,  
Hangzhou, 311100, China  
E-mail: 594283274@qq.com*

### Jiadong Wei

*Zhejiang Truly Electric Co., Ltd.,  
Hangzhou, 311100, China  
E-mail: 1050668832@qq.com*

Aiming at the problem of power quality detection and classification, a method for classifying and identifying the faults of power quality based on KPCA (Kernel principal component analysis) and FDA (Fisher Discriminant Analysis) is proposed. Using KPCA to extract the characteristics of the energy index, and the high-dimensional information of the index is deeply excavated. According to FDA, the extracted principal components are classified with high precision, and the training results are adjusted by the training array and the test array. Finally, the central eigenvector of the six types of power quality disturbance is determined, and the detected power quality data are classified. In accordance with the experimental results, the KPCA-FDA is used to classify the six types of power quality, which are more precise than PCA and KPCA in the correctness of various power quality fault classification.

*ISCC2017  
16-17 December 2017  
Guangzhou, China*

## 1. Introduction

With the continuous development of the power market, all aspects are in-depth refinement, and the tendency of the high quality power needs are getting more and more prominent. Power quality has become an important basis for the price of electricity. In this paper, the power quality disturbance results (including normal data) are divided into the following six categories: normal data, voltage fluctuations, voltage interruption, harmonic effects, voltage transient fluctuations, and oscillation transient. [1]

There are three defects in the current power quality classification method. First, most of the methods are not digging sufficiently into the link among the various indicators so to integrate them into the model. Second, many methods put stress on studying power quality indicators by weight matrix from expert assessment; Third, interference and noise to the classification results tend to have a greater impact than assumed. [2]

In view of the above drawbacks, the KPCA method is applied to increase the power quality dimension, so that the connection between the original data is clearer, and the optimal classification of the data is realized by the FDA model. At the same time, the combination of the above methods is robust to avoid the noise and interference, which is difficult to be assessed.

## 2. Analysis of KPCA Model

Compared with the general PCA method, KPCA can deeply excavate the nonlinear relationship in the data, and fully reflect the nonlinear information in the original data. The nonlinear data is projected into the high-dimensional space by the kernel function, the linear inseparable data in the low-dimensional space becomes linearly separable, and then the general PCA principle is used to reduce the dimension of the data. [3] The eigenvector of the feature is chosen according to the eigenvalue (i.e. variance) of the solution, and the number of principal components is determined to satisfy the accuracy of the data requirement.

For the sample data ( $i = 1, 2, \dots, m$ ) that has been normalized, the sample data is mapped into the high dimensional space by using the nonlinear mapping function. For the set of sample data after elevation, the matrix is:

$$C = \frac{1}{n} \sum_{i=1}^n \Phi(x_i)\Phi(x_i)^T \quad (2.1)$$

In general PCA algorithm, the characteristic equation is used,

$$\lambda\alpha = C\alpha \quad (2.2)$$

Formula is the eigenvalue of the covariance matrix and is the eigenvector with the corresponding relation. In order to ensure that most of the variance information is obtained, the variance contribution rate of the number of selected principal components  $t$  is not less than 85%.

$$\frac{\sum_{j=1}^t \lambda_j}{\sum \lambda_j} \geq 85\% \quad (2.3)$$

However, for KPCA, due to the specific form of the mapping function and the nature of the kernel function, the eigenvalue and the projection of the sample data can be obtained on the base vector in the high dimension space. The specific derivation process is as follows:

$$C = \frac{1}{N} \sum_{i=1}^n \Phi(x_i) \Phi(x_i)^T \quad (2.4)$$

Setting,

The matrix have the same eigenvalue, then the eigenvalue of is

$$u = \frac{1}{\|X^T v\|} X^T v = \frac{1}{\sqrt{\lambda}} X^T v \quad (2.5)$$

It can be summarized from the above derivation that even if the specific form of the mapping function is unknown, the eigenvalues (variance contribution rate) and the dimension data in the high dimensional space can be calculated by selecting the kernel function on the relevant eigenvector projection. The following formula gives the projection for the eigenvector of the sample data in the high dimensional space:

$$y_i = u^T \phi(x_j) \quad (2.6)$$

In this paper, the Gaussian radial basis function (RBF) kernel function is selected as the kernel function:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2.7)$$

### 3. Analysis of FDA Model

The Fisher's discriminant analysis method chooses the projection vector when the Fisher criterion function reaches the extreme value, so that the sample data meets the requirement of the intra-class spread function reaching its maximum and the between-class spread function reaching its minimum after data project in this direction. FDA is adopted to classify different types of categories of the pattern classification methods.

The FDA model is described as follows: the sample space is composed of data, and each data contains measurement variable, the data can be divided into class, the number of each data is set to (). This is to be followed by defining the intra-class dispersion matrix as the inter-class dispersion matrix. According to the data definition of the sample space, the intra-class dispersion matrix is:

$$S_a = \sum_{j=1}^p S_j = \sum_{j=1}^p \sum_{x_j \in x_j} (x_j - \bar{x}_j)(x_j - \bar{x}_j)^T \quad (3.1)$$

In the above formula, is the set of no. data for the class, is the mean vector of , and the intra-class matrix is:

$$S_b = \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (3.2)$$

In the above formula,  $\bar{w}$  is the mean vector of  $w$  and  $\bar{w}_a$  is the mean vector of all sample data. Besides, the objective function of FDA model is:

$$J = \frac{w' S_b w}{w' S_a w} \quad (3.3)$$

When the objective function reaches its maximum, that is, the intra-class spread function reaches its maximum and the between-class spread function reaches its minimum, the projection direction has the best effect on the data classification of the sample space. In order to solve this problem, the above equation is transformed into a generalized characteristic equation like the following:

$$S_b w = \lambda S_a w \quad (3.4)$$

By solving the above equation,  $n$ -dimensional eigenvalues  $\lambda$  and eigenvectors  $\alpha$  are generated, combining the eigenvectors as a feature matrix.

#### 4. Fault Detection and Type Judgment based On KPCA-FDA

For the six indicators of input power quality, each indicator is a set of isochronous time measurement data. Via using KPCA on the six input power quality data, the original input data calculates the principal component of the high-dimension data through the RBF kernel function elevation to high dimensional space. The principal component types can be greater than six, in comparison with the normal PCA method, the maximum extraction of the original data can be contained into the principal component, which can effectively enhance the accuracy of fault detection. Given the number of principal components extracted by KPCA, the model is generated by the FDA from the principal component, and the feature matrix is obtained. Compared with the way of determining the fault type through analysis from  $\xi$ , or other passing and statistic, using KPCA-FDA to get the high-dimension data and similarity function of the projection vector by the feature matrix  $\alpha$ , then analyzing statistics on distance and the nuclear density estimation to determine the fault type, it is obviously that more accurate diagnosis results can be delivered, with the latter method.

Under normal circumstances, the greater the difference between the input sample data, the better effect on the training classifier. Setting the sample data of the classifier respectively into training data vector  $\xi$  and test data vector  $\xi'$ , may contribute to the following formula as fault classification basis:

$$\xi = \Gamma \cdot F_n \quad (4.1)$$

In the above formula, the row vector  $\xi$  or  $\xi'$  are composed of the former principal component extracted by KPCA.  $\Gamma$  is the characteristic matrix calculated by the FDA modeling.

Each type of training data is categorized into a fault group and the mean of every training data vector group is found, to be followed by defining the Euclidean distance between the training data discriminant vector  $\xi_p$  and the training data mean vector  $\xi_{center,p}$  as

$$Dis_p = \left\| \xi_{train,p} - \xi_{center,p} \right\| \quad (4.2)$$

The Parzen window method can be applied to re-determine the statistical Euclidean distance indicator. And the after-debugging Euclidean distance is defined

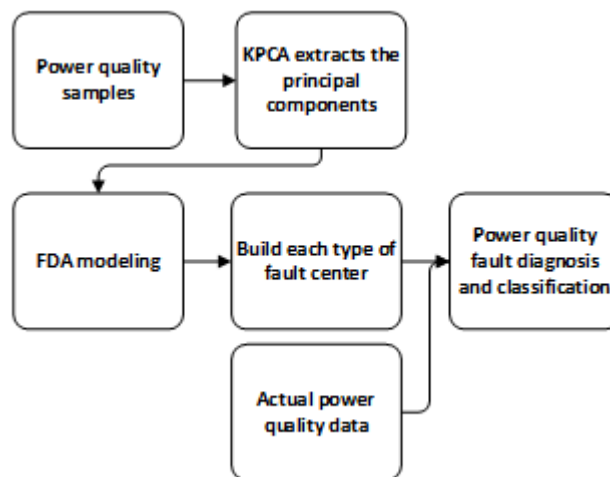
, as the condition for determining whether a fault occurs in the test data, and when the test data is taken as an input, the calculated distance can be obtained as a fault detection criterion through the above calculation process, that is,

$$\xi'_p = \left\| \xi_{text} - \xi_{center,p} \right\| \tag{4.3}$$

If is greater than, it can be judged as the type of fault occurrence. Through traversing all types of fault, the index of the fault diagnosis of power quality and the type of faults can be obtained. In accordance with the general methods, only, or other passing and statistic are adopted to for classification, while the method in this paper, after being processed by KPCA-FDA, is to get the principal components of the data after dimension up-scaling and the characteristic matrix of Fisher discriminant analysis model through the process of are obtained. The Parzen window method is employed to determine the classification basis, effectively using the implicit information of the sample data, and improving the degree of separation among the various fault categories. The KPCA-FDA method effectively improves the power quality indicators of detection, classification and classification capabilities.

Detection and classification process of KPCA-FDA method:

- (1) Collecting six types of energy index data set and 100 sets of historical power quality data set .
- (2) If the contribution rate of variance is greater than 85%, the number of principal components n shall be determined and KPCA shall be used to extract the core principal components and .
- (3) Using the FDA modeling to reprogram the optimal projection direction, with the result of the characteristic matrix being.
- (4) Calculating the training data and of each class and finally determining the fault detection criteria.
- (5) Collecting the power quality data online, and calculating the Euclidean distance and each type's Euclidean distance.
- (6) Comparing with the corresponding fault detection criteria, if there is a class of fault occurrence.



**Figure 1:** KPCA-FDA method flow chart

## 5. Instance Analysis

The impact of power quality indicators are divided into the following categories: voltage deviation, frequency deviation, three-phase voltage unbalance, voltage fluctuations and flicker, total harmonic distortion, voltage sag and interruption. The power quality faults set to six categories: normal data, voltage fluctuations, voltage interruptions, harmonic effects, shock transient, and temporary voltage fluctuations. In order to validate the effectiveness of the method proposed in this paper, 300 sets of training data are randomly generated by matlab. Each type of power quality fault contains 50 groups, and 180 groups of test data (30 in each group) are generated at the same time to verify the classification recognition effect. By setting sampling frequency to 6.4 kHz, 1,024 sampling points take eight cycles of waveform data with the duration of 0.16s.

The above 300 groups of training data are input for KPCA to extract the principal components. If the number of principal components is 21, the variance contribution rate will be more than 85%. When the number of principal components is 23, the recognition rate of the fault classification will reach 95%. Considering the comprehensive calculation speed and recognition accuracy, the number of main components is determined as 21. For 300 sets of training data set, RBF kernel function is used for processing, and. According to the eigenvalues (variance contribution rate), the first 21 principal components are extracted from the descending order, and the FDA model is calculated. Based on the criteria set for fault diagnosis of the six types of power quality, the correctness of the classification results is compared with that of the traditional methods as shown in Table 1:

Types	PCA	KPCA	KPCA-FDA
Normal data	98%	100%	100%
Voltage fluctuations	85%	89%	94%
Voltage interruption	90%	94%	100%
Harmonic effects	87%	95%	99%
Temporary voltage fluctuations	86%	93%	97%
Shock transient	82%	83%	94%
Comprehensive recognition rate	88%	92%	97%

**Table 1:** different methods of recognition rate

As can be seen from Table 1, the KPCA-FDA method introduced in this paper has a remarkable improvement, in terms of voltage fluctuation and transient oscillation with a correct recognition rate of 94%. While the correct recognition rate of voltage interruption and normal data are even raised to 100 %, the influence of harmonics and temporary voltage fluctuation has been greatly improved. The KPCA-FDA method is obviously superior to the PCA and KPCA methods concerning comprehensive recognition rate and various types of fault recognition rate.

## 6. Conclusion

Compared with the traditional PCA and KPCA methods, the proposed KPCA-FDA classification method effectively integrates the advantages of the two above methods. At the same time, the experimental results also prove that KPCA methods are used to extract six types of power quality index information, and then through the FDA model to achieve the best classification. Finally, coordinating with the results of KPCA-FDA, the method's feasibility and

superiority of the fault diagnosis of power quality and classification has been provided, which inspires a new idea of the fault diagnosis of power quality and classification.

## References

- [1] Zhan Yong, Cheng Haozhong, Ding Yifeng. *Classification and identification of power quality disturbances based on S-transform and support vector machine* [J], Proceedings of the CSEE, 2005,25 (4): 51-56. In Chinese
- [2] ZHAO Xia, ZHAO Chengyong, JIA Xiufang, LI Gengyin. *Fuzzy Comprehensive Evaluation of Power Quality Based on Variable Weights* [J]. Power System Technology, 2005, 29 (6): 11 -16 In Chinese
- [3] Wei Zhenzhong, *Feature Extraction Method Based on Principal Component Analysis* [J] Journal of Guangxi University of Technology, 2006, 17 (4): 27-31. In Chinese