

Power Quality Disturbances Identification Method Based on Novel Hybrid Kernel Function

Liquan Zhao* and Meijiao Gai*

Abstract

A hybrid kernel function of support vector machine is proposed to improve the classification performance of power quality disturbances. The kernel function mathematical model of support vector machine directly affects the classification performance. Different types of kernel functions have different generalization ability and learning ability. The single kernel function cannot have better ability both in learning and generalization. To overcome this problem, we propose a hybrid kernel function that is composed of two single kernel functions to improve both the ability in generation and learning. In simulations, we respectively used the single and multiple power quality disturbances to test classification performance of support vector machine algorithm with the proposed hybrid kernel function. Compared with other support vector machine algorithms, the improved support vector machine algorithm has better performance for the classification of power quality signals with single and multiple disturbances.

Keywords

Hybrid Kernel Function, Power Quality Disturbance, Support Vector Machine, Wavelet Transform

1. Introduction

The expansion of power system scale implies the augmentation of non-linear and impact loads, power quality disturbances (PQDs) problems have become increasingly prominent [1,2]. To ensure the safe and stable operation of the power system, the power quality must be detected and analyzed. For this purpose, it is imperative to identify the potential PQDs. The PQDs mainly contain single disturbance and multiple disturbances. The single PQDs mainly include harmonic, transient oscillation, flicker, voltage interruption, etc., and the typical multiple PQDs include harmonic with voltage sag or voltage swell, flicker with voltage swell or voltage sag, etc.

The classification processes of PQDs are composed of feature extraction and classification. In feature extraction process, we use wavelet transform (WT) to decompose the observed PQDs signal. If we directly use the transformation coefficient as feature vector, this will cause the system to learn slowly and the structure is huge. Moreover, the transient energy distribution of the signal after WT is similar and difficult to distinguish. Therefore, we use the normalized transient wavelet energy difference of the disturbance

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received September 7, 2017; first revision February 19, 2018; accepted March 12, 2018.

Corresponding Author: Liquan Zhao (zhao_liquan@163.com)

* Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education (Northeast Electric Power University), Jilin, China (zhao_liquan@163.com, 516104653@qq.com)

signal and standard signal as the feature vector. Eventually, we import the feature vector into the classifier for classification. The feature vector is more conducive by this method. In the classification process, we use the support vector machine (SVM) to classify the feature vectors that are obtained by the WT method. The kernel function mathematical model of SVM directly affects the classification performance. Different types of kernel function have different generalization ability and learning ability. The single kernel function cannot have better ability both in learning and generalization. Therefore, we propose a hybrid kernel function that is composed of two single kernel functions to improve both the ability in generation and learning. We also obtain the optimal combination coefficient of the two kernel functions through multiple experiments.

2. Related Work

The identification of the PQDs has two processes: feature extraction and classification. The feature extraction methods are typically based on the short time Fourier transform (STFT), Fourier transform [2], wavelet transform [3], and S transform (ST) [4,5]. The Fourier transform, which is most suitable for stationary signals, provides information about the frequency components, but does not reflect when the signal exists and for how long. Most PQDs are non-stationary, thus we need to determine the frequency and the time of occurrence of the disturbance signals. The short time Fourier transform generates information about the frequency as well as the time. However, the width of the window is a constant; hence it cannot accurately describe the characteristics of the disturbance. Different disturbance needs different window width, which poses difficulties in choosing the window width for the STFT. At the same time, the STFT cannot trace the transient and mutation signals. These drawbacks can be overcome by the waveform transform. The WT has higher time resolution and frequency resolution for the low frequency signal and high frequency signal respectively [3]. Consequently, this method can satisfy the requirement of the resolution for classification of the PQD signal. The signal adaptability of the WT makes it precise, meaning it can successfully be used to analyze the non-stationary signals. However, the WT is sensitive to noise. The ST can be viewed as an extension to the WT and the STFT. The ST is a form of multi-resolution analysis, which can provide information about the frequency and the phase [5]. However, the resulting ST matrix is redundant. Thus, the computing time of ST is long.

The commonly used classification methods are artificial neural networks (ANN) [6], decision tree [7] and SVM [8-10]. The ANN methods are simple in structure, strong in problem solving, and are defined by large-scale distributed parallel processing, nonlinearity, self-organization and self-learning. However, the algorithm has some disadvantages, such as the local optimization problems and poor convergence. In addition, the training time of this method can be long and over-fitting may occur. Decision tree is meant to simulate human thought, however, the regulation of a decision tree is hard to establish. The SVM methods can effectively solve the nonlinear, finite sample and high dimensional pattern recognition problems. The basic principle of SVM methods is to transform the feature vector that is inseparable in the low-dimensional space to the high-dimensional space and make it be separable in the high-dimensional space. Different kernel functions determine different SVM classifiers. The kernel function mathematical model of SVM directly affects the performance of classification, so it is very important to design a suitable kernel function [11].

3. Feature Extraction Using the WT

The WT has a variable time-frequency window, which can be adjusted according to the difference in signal frequency. Furthermore, the WT has good time-frequency local characteristics [12]. We can obtain a series of coefficients corresponding to each scale by multi-scale decomposition of the disturbance signal. These coefficients are the basis of feature extraction using the WT. The input energy is loaded into the wavelet coefficients, as follows:

$$\int [x(t)]^2 dt = \sum_{n=1}^N |C_j(n)|^2 + \sum_{x=1}^j \sum_{n=1}^N |D_x(n)|^2 \quad (1)$$

where, $x(t)$ is the signal to be decomposed, $C_j(n)$ and $D_x(n)$ are respectively the approximation coefficient and detail coefficient of wavelet decomposition. The approximate coefficients store the fundamental wave energy, and the detail coefficients store the transient energy. As the PQD occurs, each band energy of the signal changes. Moreover, the transient energy of different PQD signal differs in different frequency bands.

As the Daubechies wavelets (db) offer tight support, are sensitive to irregular signals, and generate an orthogonal analysis, we used the db4 wavelet to decompose the disturbance signal for 8 layers. The relationship between the wavelet coefficients and wavelet energy can be expressed as follows:

$$E_{d_j} = \sum_n (d_j(n))^2 \quad (2)$$

where, E_{d_j} and $d_j(n)$ are respectively the transient energy and detail coefficient of the j^{th} layer. We used the waveform transform method to decompose the signal into 8 layers, and obtained the transient energy E_i of the signal. The relatively small differences in transient energy distribution of each disturbance negatively impacted the classification. Consequently, we used transient energy difference between the disturbance signal E_i and standard signal E_{ref} to construct feature vector X , which is viewed as the subsequent input vector of the SVM.

$$X = [E_1^*, E_2^*, E_3^*, E_4^*, E_5^*, E_6^*, E_7^*, E_8^*] \quad (3)$$

where $E_i^* = E_i - E_{ref}$.

4. Support Vector Machine

The main principle of SVM classification is that transforming the inseparable data in low-dimensional space to high-dimensional separable space by utilizing a kernel function. The mapped data can be separated in the new space. Moreover, the SVM designs a classification hyper-plane as a decision plane to realize the classification. Since the optimal solution of the SVM is based on the idea of minimizing structural risk, it has stronger generalization ability than the method of nonlinear-function approximation.

We set the training set as $D = \{(x_i, y_j), x_i \in R^n, y_j \in \{+1, -1\}, i = 1, 2, \dots, N\}$, where x_i and $y_j \in \{+1, -1\}$ are the training vector set and label vector set, respectively. The function of the optimal separating hyper-plane is expressed as follows:

$$d(x) = w^T x + b = 0 \quad w \in R^n, b \in R \tag{4}$$

When the y_i can meet the following condition

$$y_i(w^T x_i + b) >= 1 \quad i = 1, 2, \dots, N \tag{5}$$

The classification hyper-plane can makes the classification interval maximum and successfully separate the feature vectors. The maximum classification interval is

$$d = \frac{2}{\|w\|} \tag{6}$$

A bigger the maximum classification interval means that the SVM will have a better performance of classification. Therefore, maximizing the classification interval can be converted into mining $\|w\|$. There, the cost function is expressed as:

$$\begin{cases} \min \frac{1}{2} w^T w \\ st \ y_i(w^T x + b) - 1 \geq 0, \quad i = 1, 2, \dots, N \end{cases} \tag{7}$$

SVM algorithm uses the Lagrange method to solve the above problem. The Lagrange function is:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [y_i(w^T x + b) - 1] \tag{8}$$

where, the Lagrange multiplier α_i is a non-negative value. The derivatives of w and b in (8) are expressed as:

$$\begin{cases} w = \sum_{i=1}^N \alpha_i y_i x_i \\ \sum_{i=1}^N y_i \alpha_i = 0 \end{cases} \tag{9}$$

The Lagrange function in (8) is transformed into the problem about α by introducing (9) into (8). That is

$$\begin{aligned} & \max_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ & st \ \alpha_i \geq 0, \quad i = 1, 2, \dots, N \\ & \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \tag{10}$$

The optimal solution can be obtained by using sequential minimal optimization method, and expressed as α^* . Based on the optimal solution α^* , we can get the optimal solutions of w and b .

$$w^* = \sum_{i=1}^N \alpha_i^* y_i x_i \tag{11}$$

$$b^* = y_i - \sum_{i=1}^N \alpha_i^* y_i(x_i x_j) \tag{12}$$

The training set is named support vector when the optimal solution α^* is not zero. Therefore, the classification function is expressed as

$$d(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* y_i(x_i x_j) + b^*) \tag{13}$$

where, $\text{sgn}()$ is a sign function, n is the number of the support vector. The nonlinear function $K(x_i, x_j)$ named kernel function in SVM is used to replace (x_i, x_j) in (13) to transform the feature vectors from inseparable space to separable space. Therefore, the classification function mathematical expression based on kernel function is as follows:

$$d(x) = \text{sgn}\left(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b^*\right) \quad (14)$$

5. Improved SVM Based on a Hybrid Kernel Function

The kernel function mathematical model of SVM directly affects the classification performance. Different types of kernel functions have different generalization ability and learning ability. The kernel function transforms the inseparable data in low-dimensional space to high-dimensional separable space. At the same time, it does not increase the computation complexity and running time. The typical kernel functions are as follows:

(1) Gaussian kernel

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / p^2) \quad (15)$$

(2) Polynomial kernel

$$K(x_i, x_j) = ((x_i \cdot x_j) + 1)^d \quad (16)$$

(3) Sigmoid kernel

$$K(x_i, x_j) = \tanh(v(x_i \cdot x_j + r)) \quad (17)$$

where p , d , v , and r are real constants.

Different types of kernel functions have different generalization ability and learning ability. Fig. 1 shows the curve of the radial basis function (RBF), when the parameter p respectively equals 0.1, 0.2, 0.3, 0.4, 0.5, with 0.2 as the test point. As shown in Fig. 1, when the input data were near the test point, the value of the kernel changed significantly, indicating that the Gaussian kernel function is a child of the local kernel functions.

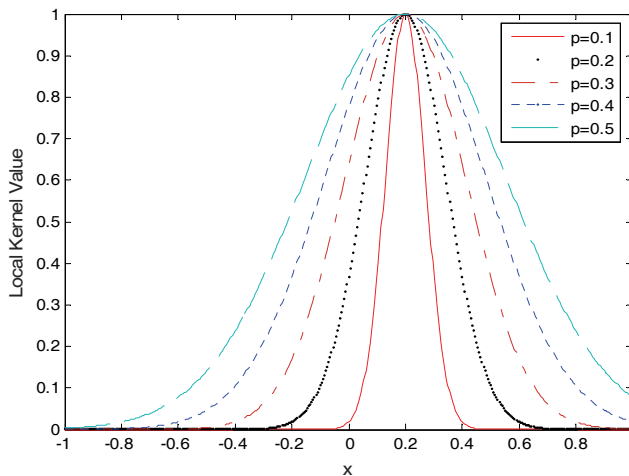


Fig. 1. Radial basis kernel function.

Fig. 2 shows the curve of the polynomial kernel, when the parameter d respectively equals 1, 2, 3, 4, 5, with 0.2 as the test point. As evidenced by Fig. 2, when the input data were far from the test point, the value of the kernel changed significantly, indicating that the polynomial kernel is the global kernel function.

The K_1 , K_2 are kernel functions, set λ is constant and $\lambda \geq 0$, the kernel functions are constructed according to the following formulas

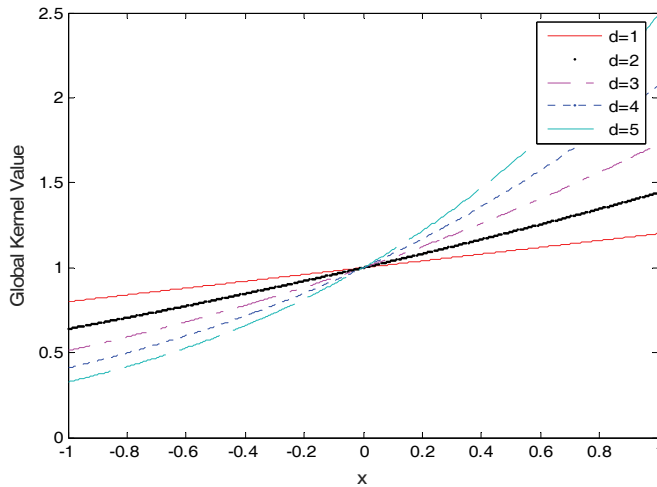


Fig. 2. Polynomial kernel function.

$$K(x_i, x_j) = K_1(x_i, x_j) + K_2(x_i, x_j) \tag{18}$$

$$K(x_i, x_j) = \lambda K_1(x_i, x_j) \tag{19}$$

$$K(x_i, x_j) = K_1(x_i, x_j) \cdot K_2(x_i, x_j) \tag{20}$$

The hybrid kernel function, which is constructed by simple, single kernel functions, still satisfies Mercer’s theorem of the kernel function.

For K_1, K_2 defined on a limited set of points $\{x_1, K, x_n\}$ for any vector $\alpha \in R^n$, K is a positive, semi-definite matrix, for which all α must satisfy $\alpha'K\alpha \geq 0$ as the necessary and sufficient condition. $\alpha'(K_1 + K_2)\alpha = \alpha'K_1\alpha + \alpha'K_2\alpha \geq 0$, where the sum of K_1 and K_2 is positive semi-definite matrix, and the kernel function is $K(x, z) = K_1(x, z) + K_2(x, z)$. Therefore, it can conclude that it is still a kernel function for $K = \lambda K_1 + (1 - \lambda)K_2$, $\lambda \in (0, 1)$.

Although the polynomial kernel and the RBF kernel are typical kernel functions of SVM, both of them have their own limitations. Any single kernel function cannot have better ability both in learning and generalization. To overcome the problem, we use two kernel functions to construct a new mixed kernel function that is suitable for the classification of PQDs. According to the constitute conditions of the kernel function, the sum of two kernel functions is still a kernel function. Therefore, we mixed the polynomial kernel and the RBF kernel to construct a new kernel function of SVM to make the SVM with the new kernel function have better performance in classification of PQDs. The proposed kernel function of SVM can be expressed as:

$$K(x_i, x_j) = \lambda_1 \exp(-\|x_i - x_j\|^2 / p^2) + \lambda_2 ((x_i \cdot x_j) + 1)^d \tag{21}$$

where λ_1, λ_2 are the proportionality coefficients of the two kernel functions, their ranges are from zero to one, and their sum is one.

Figs. 3 and 4 are the output diagrams of the proposed kernel function with different test points. As evidenced by Figs. 3 and 4, the new kernel function not only highlights the local characteristics of data near the test point, but also retains the global characteristics of data far from the test point. Through a series of adjustments of the coefficients, we found that, when $\lambda_1 = \lambda_2 = 0.5$, the proposed kernel function has better performance.

The procedure of the proposed method is as follows:

- 1) Using the wavelet transform to decompose the PQD signal
- 2) Extracting the wavelet energy difference between the disturbance and standard signal
- 3) Normalizing the wavelet energy difference, and using it as the feature vector
- 4) Identifying the disturbance signals using the SVM based on the proposed hybrid kernel function

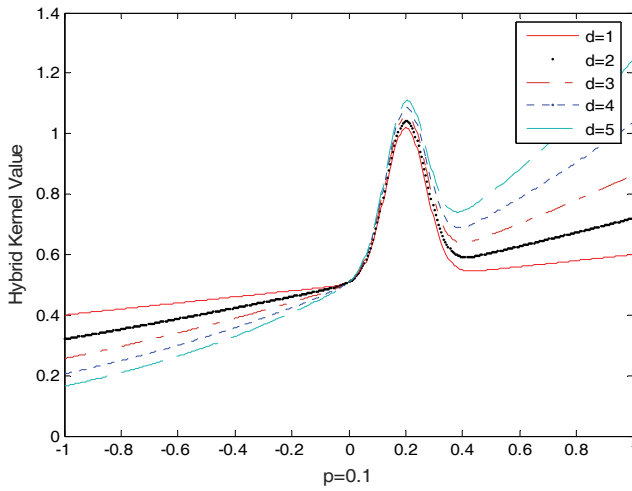


Fig. 3. Hybrid kernel function with $p=0.1$.

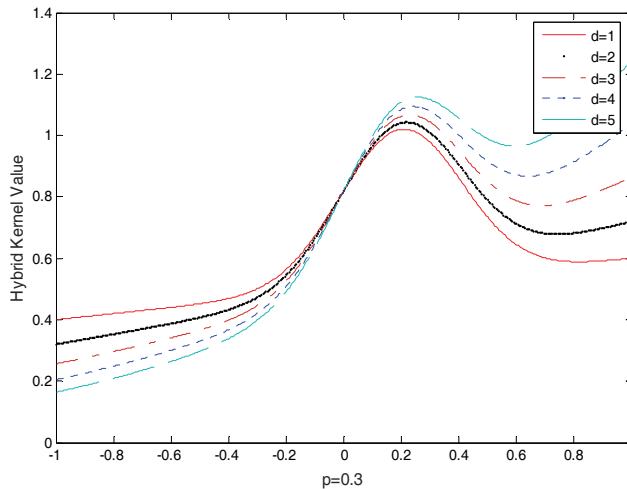


Fig. 4. Hybrid kernel function with $p=0.3$.

6. Simulation and Analysis

In this paper, we chose seven types of single PQDs (i.e., harmonic, voltage sag, flicker, voltage swell, transient oscillation, transient pulse, and voltage interruption) and four types of multiple PQDs (i.e., flicker/harmonic with voltage swell/voltage sag) to simulate. For each disturbance signal, 200 samples were selected. The mathematical model is shown in reference [12]. The fundamental frequency of the signal was 50HZ. There were 1281 samples points per cycle. The length of the signal is ten cycles. In order to simulate the real PQD signal, we add 15db of Gaussian white noise to each disturbance signal. As already mentioned, the process of PQD identification contains feature extraction and classification. Firstly, the PQD signal is decomposed for 8 layers by using the db4 wavelet as the mother wavelet, then, the wavelet coefficients of each layer were extracted. It is observed that the trends of the wavelet coefficients of the signals was roughly the same and difficult to distinguish. Based on Parseval's theorem, we firstly compute the energy differences between the standard signal and PQDs signal, and use the obtained eight energy differences as the characteristic vector. Secondly, we normalize the vector to construct the feature vector. In the end, SVM algorithms with different kernel functions are used to classify the constructed feature vector.

The number of samples are 200, and 80 are selected as training samples to obtain the mathematical model of the decision function in the SVM algorithm. We use the remaining 120 samples as test samples to test the classification accuracy. The classification accuracies of the eleven PQD signals for the SVM classifier with the polynomial kernel, the proposed hybrid kernel function, the RBF kernel and the other method in [11,13] are shown in Table 1.

From Table 1, it can be seen that classification accuracy of SVM algorithm with proposed hybrid kernel is higher than the other classification methods. For voltage swell, the proposed SVM algorithm has the same classification accuracy with the two methods in [11,13], which is higher than the SVM algorithms based on polynomial kernel and RBF kernel. For voltage interruption and voltage sag, the proposed method has lower accuracy than the methods in [11,13], which is higher than the others. For harmonic and transient oscillation, all the methods are the same, as the accuracy is 100%. For transient pulse, voltage flicker, voltage swell with harmonic and voltage sag with harmonic, the classification accuracy of improved SVM algorithm is higher than the other SVM algorithms. For voltage sag with harmonic, the classification accuracy of proposed SVM method is higher than the other classification methods except for the SVM algorithm with radial basis kernel function. For voltage swell with flicker, the proposed method has the same accuracy with the method in [11], which is higher than the other three methods. For voltage sag with flicker, the classification accuracy of proposed SVM algorithm is lower than the other SVM algorithm. However, the average accuracy of the SVM algorithm with the proposed hybrid kernel function is higher than other algorithms. The classification accuracies of voltage sag and voltage interruption are both not high. The amplitude is also reduced in the case of voltage interruption, which is different from the sag. This suggests that there is a certain degree of interference in the PQDs classification, meaning that the proposed SVM method has some limitations. Future work should be aimed at improving the classification threshold, which would, in turn, improve the overall classification performance of the proposed SVM algorithm. We mainly focus on SVM algorithm so we did not compare the other intelligence algorithms [14-16] that can be used to optimize the SVM and classify quality disturbances.

Table 1. Classification accuracies of the single and multiple power quality disturbances

Disturbance type	Classification accuracy (%)				Hybrid kernel function
	Radial basis kernel function	Polynomial kernel function	Method [13]	Method [11]	
Voltage swell	99.1667	98.3333	100	100	100
Voltage sag	70	84.1667	93.3333	99.1667	85
Voltage interruption	80	72.5	94.1667	96.6667	81.1667
Harmonic	100	100	100	100	100
Transient pulse	96.6667	97.5	80.8333	70.6667	99.1667
Transient oscillation	100	100	100	100	100
Voltage flicker	99.1667	100	83.1667	89.1667	100
Voltage swell with harmonic	98.3333	100	100	99.1667	100
Voltage sag with harmonic	99.1667	97.5	97.5	96.1667	98.3333
Voltage swell with flicker	97.5	97.5	98.3333	99.1667	99.1667
Voltage sag with flicker	98.3333	95.8333	99.1667	100	95
Average classification accuracy	94.3939	94.8485	95.13636364	95.46971818	96.2121

The average classification accuracies that are obtained by using the SVM algorithms with different kernel functions are shown in Fig. 5. From this figure, we can see that the average classification accuracy that is obtained by using the SVM algorithm with the proposed kernel function is higher than the SVM algorithms with the other kernel function.

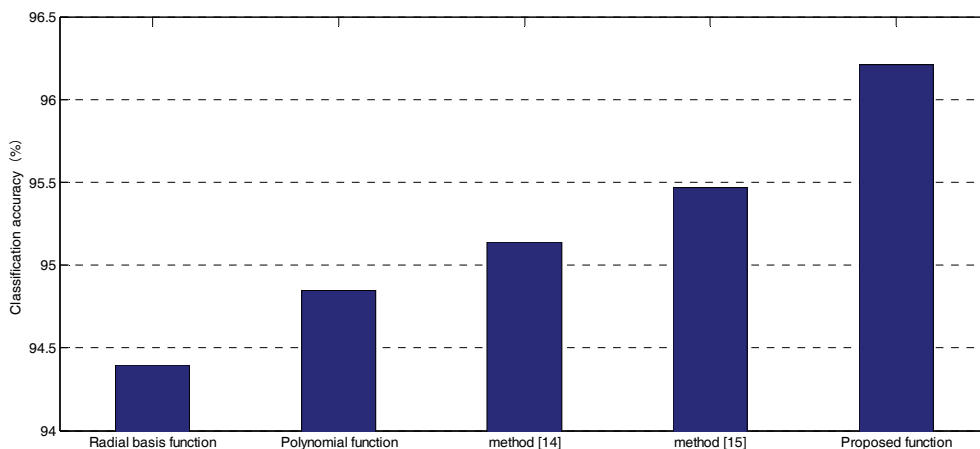


Fig. 5. Comparison of average classification accuracy for different methods.

7. Conclusion

In feature extraction, this paper used the wavelet energy difference between the standard signal and the PQD signal to construct the feature vector. In feature vector classification, this paper used the improved SVM with proposed hybrid kernel function to improve the classification accuracy of PQDs. The proposed

SVM method has higher generalization and learning abilities than the others, and its classification accuracy is greatly improved.

Acknowledgement

This paper is supported by the Foundation of Jilin Educational Committee (Grant No. 2015235).

References

- [1] W. G. Morsi and M. E. El-Hawary, "Power quality evaluation in smart grids considering modern distortion in electric power systems," *Electric Power Systems Research*, vol. 81, no. 5, pp. 1117-1123, 2011.
- [2] Z. Liu, Q. Zhang, and Y. Zhang, "Review of power quality mixed disturbances identification," *Power System Protection and Control*, vol. 41, no. 13, pp. 146-153, 2013.
- [3] D. De Yong, S. Bhowmik, and F. Magnago, "An effective power quality classifier using wavelet transform and support vector machines," *Expert Systems with Applications*, vol. 42, no. 15-16, pp. 6075-6081, 2015.
- [4] Y. Wu, Q. Tang, Z. Teng, N. Li, and X. Wang, "Feature extraction method of power quality disturbance signals based on modified S-transform," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 36, no. 10, pp. 2682-2689, 2016.
- [5] F. Xu, H. Yang, M. Ye, Y. Liu, and J. Hui, "Classification for power quality short duration disturbances based on generalized S-transform," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 32, no. 4, pp. 77-84, 2012.
- [6] Y. Wang, Y. Li, Z. Qu, and S. Liu, "The classification of power quality disturbance based on PSO-MP algorithm and RBF neural network," *Electrical Measurement & Instrumentations*, vol. 53, no. 13, pp. 54-58, 2016.
- [7] H. Chen and G. Zhang, "Power quality disturbance identification based on decision tree and support vector machine," *Power Grid Technology*, vol. 37, no. 5, pp. 1272-1278, 2013.
- [8] G. Han and X. Chu, "Power quality disturbance classification based on multi-features combination and optimizing parameters of SVM," *Proceedings of the CSU-EPSA*, vol. 27, no. 8, pp. 71-76, 2015.
- [9] X. Yang, B. Sun, X. Zhang, and L. Li, "Short-term wind speed forecasting based on support vector machine with similar data," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 32, no. 4, pp. 35-41, 2012.
- [10] Z. Liu, Y. Cui, and W. Li, "A classification method for complex power quality disturbances using EEMD and rank wavelet SVM," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1678-1685, 2015.
- [11] G. Liu and X. Yang, "Support vector machine with mixed kernel function," *Microcomputer & Applications*, vol. 36, no. 11, pp. 19-22, 2017.
- [12] L. Zhao and Y. Long, "Classification of power quality composite disturbance based on improved SVM," *Advanced Technology of Electrical Engineering and Energy*, vol. 35 no. 10, pp. 63-68, 2016.
- [13] Z. Chen, M. Ouyang, and H. Liu, "PQD recognition based on wavelet energy difference distribution and SVM," *Computer Engineering and Applications*, vol. 47, no. 20, pp. 241-244, 2011.
- [14] D. Wang, X. Wu, and Z. Wang, "Fault location for distribution network with distributed power based on improved genetic algorithm," *Journal of Northeast Dianli University (Natural Science Edition)*, vol. 36, no. 1, pp. 1-7, 2016.
- [15] F. J. Kuang and S. Y. Zhang, "A novel network intrusion detection based on support vector machine and tent chaos artificial bee colony algorithm," *Journal of Network Intelligence*, vol. 36, no. 1, pp. 1-7, 2016.

- [16] W. Qian, L. Lina, X. Bei, and C. Houhe, "Planning of distributed energy storage system for improving low voltage and network loss in rural network," *Journal of Northeast Electric Power University*, vol. 37, no. 5, pp. 19-24, 2017.



Liquan Zhao <https://orcid.org/0000-0002-9499-1911>

He is an assistant professor at Northeast Electric Power University, China. He received his Ph.D. degree in college of information and communication from Harbin Engineering University in 2009. His current research interests include wireless sensor networks and blind source separation.



Meijiao Gai <https://orcid.org/0000-0002-9143-1863>

She received her Bachelor's degree at Baicheng Normal University in 2012. She is now a M.S. student in college of information engineering at Northeast Electric Power University, Jilin. She is interest in the recognition of multiple power quality disturbance.