Journal of Information Hiding and Multimedia Signal Processing ©2019 ISSN 2073-4212
Ubiquitous International Volume **10**, Number **1**, January **2019**

Classification of Multiple Power Quality Disturbances Based on PSO-SVM of Hybrid Kernel Function

Liquan Zhao[1], Meijiao Gai[2], Yanfei Jia[3]

[1]College of Information Engineering
Northeast Electric Power University
NO.169 Changchun Road,Chuanying District, Jilin City, China
[2]Telecommunication Branch
Qinzhou Power Supply Bureau, China
[3]College of Electrical and Information Engineering
Beihua University, China
zhao_liquan@163.com

Received July 2018; revised September 2018

Abstract. *To improve the classification accuracy of multiple power quality disturbances, we proposed a hybrid kernel function support vector machine (HSVM) based on the particle swarm optimization algorithm for classification. The support vector machine based on hybrid kernel function has better learning ability and generalization ability compared with single kernel function. At the same time, the nuclear parameter and the weight coefficient of single kernel function in the hybrid kernel function have no effective standard. Therefore, we use the particle swarm algorithm to optimize the parameters of the HSVM, ultimately determining the optimal parameters. We use the optimized HSVM to classify for seven single power quality disturbance and four multiple power quality disturbances. The experiment results show that the proposed method has higher accuracy in its classification performance compared with the traditional method.*
Keywords: Power quality disturbances; Support vector machine; Hybrid kernel function;particle swarm optimization.

1. **Introduction.** The use of new types of electrical equipment, large-scale integration of new energy, a large number of non-linear impacting and volatile electrical equipment connecting to the grid significantly affects power quality. With the advent of more precise and sensitive high tech products, there are higher requirements for power quality. The classification of power quality disturbance is not only an important link in detecting power quality but also an important index to evaluate power quality [1].

In the electronic power system, there are not only single disturbances but also multiple disturbances. The single power quality disturbances mainly include: Voltage sag, voltage swell, voltage interruption, flicker, harmonic, transient oscillation, transient pulse, etc. The typical multiple power quality disturbances include: voltage sag with harmonic, voltage swell with harmonic, flicker with voltage sag, flicker with voltage swell, etc. The process of power quality disturbance classification has two parts: feature extraction and classification. Although there are many feature extraction and classification methods for different applications[2-6],they are not suitable for classification of power quality disturbance. The main feature extraction methods that are used for power quality disturbance classification include: Fourier transform (FT), short time Fourier transform [7](STFT), wavelet transform[8-9] (WT), and S transform[10-12] (ST). The main classification methods are as follows: neural network [13], decision tree[14] and support vector machine[15-18] (SVM). Neural network is a kind of widely used classification method that has a simple structure and strong problem solving ability. But this method also has a long training time, poor convergence and tendency to fall into local optimal problems. Decision tree is a classification algorithm based on the data structure. It constructs the classification rules by simulating human thinking. However, the process of establishing rules is very complex. Support vector machine (SVM) is developed on the basis of statistical learning, and has been widely used in the field of classification recognition.

In this paper, WT was used for feature extraction, and the improved SVM was used for classification. As the selection of nuclear parameter and weight efficient in the hybrid kernel function has an important effect on the classification performance, we adopted the PSO to optimize nuclear parameter and weight efficient in the hybrid kernel function to determine the optimal parameter. Finally, we obtained the global optimal value.

2. **Support Vector Machine.** The main idea of SVM classification is to transform data in the input space into a high-dimensional space by the nonlinear transformation that is defined by the kernel function. We then find a classification hyper-plane as a decision plane in this space to maximize the isolation edge between the positive and negative cases. The algorithm is eventually transformed into a quadratic optimization problem, and the global optimal solution is obtained. Since the optimal solution of the support vector machine is based on the idea of minimizing structural risk, it has stronger generalization ability than the method of non-linear-function approximation.The training set is $D = \{(x_i, y), x_i \in R^n, y \in \{+1, -1\}, i = 1, 2, ..., N\}$,$x_i$ denotes the training vector set and y denotes the label vector. The function of optimal separating hyper-plane is defined as

$$d(x) = w^T x + b = 0 \qquad w \in R^n, b \in R \tag{1}$$

The classification hyper-plane can accurately separate the training vector set and maximize the classification interval. Therefore, the condition which the classification should meet is

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

A maximum distance of positive and negative samples from the classification hyper-plane, namely the classification interval, is

$$d = \frac{\left|w^T x + b\right|}{\|w\|} \tag{3}$$

Therefore, constructing the optimal classification hyper-plane is converted to solve the maximum of the classification interval, namely the minimization problem of $\|w\|$

$$\left\{ \begin{array}{l} \min \ \frac{1}{2} w^T w \\ st \ \ y_i(w^T x + b) - 1 \geq 0 \end{array} \right. \qquad i = 1, 2, ...N \tag{4}$$

Using the constraint condition extremum method, the problem leads to the Lagrange function:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i \left[y_i(w^T x + b) - 1 \right] \tag{5}$$

Where, $\alpha_i \geq 0$ is non-negative Lagrange multiplier, the conclusion of the optimization problem is given by the saddle point of the Lagrange function and meets the partial derivatives of w, b.

$$\begin{cases} w = \sum\limits_{i=1}^{N} \alpha_i y_i x_i \\ \sum\limits_{i=1}^{N} y_i \alpha_i = 0 \end{cases} \tag{6}$$

The minimum is obtained by substituting the values of w and b into (5), and at the same time, we can form the dual problem:

$$\begin{array}{l} \max\limits_{\alpha} \sum\limits_{i=1}^{N} \alpha_i - \frac{1}{2}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ st \quad \alpha_i \geq 0, \qquad i = 1, 2, ...N \\ \quad \sum\limits_{i=1}^{N} \alpha_i y_i = 0 \end{array} \tag{7}$$

If α^* is the optimal solution, then

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i \tag{8}$$

$$b^* = y_i - \sum_{i=1}^{N} \alpha_i^* y_i (x_i x_j) \tag{9}$$

When $\alpha^* \neq 0$, the training is called the support vector. To sum up, the discriminated function is

$$d(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i (x_i x_j) + b^*) \tag{10}$$

Where, sgn() is the sign function, b^* is the classification threshold, and n is the number of the support vector. By using the kernel function $K(x_i, x_j)$ to replace the inner product(x_i, x_j), all of the calculations are performed in the original input space dimension. At this point, the optimal target function is defined as:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i x_j) \tag{11}$$

The kernel classification function is expressed as

$$d(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i K(x_i x_j) + b^*) \tag{12}$$

However, if in the case that linear non-separable, the slack variable ξ_i and error penalty C are introduced, then the problem of the optimal classification hyper-plane can be described as:

$$\begin{cases} \min \ \frac{1}{2} \|w\|^2 + C \sum\limits_{i=1}^{n} \xi_i \\ st \ \ y_i(w^T \phi(x_i) + b) - 1 \geq 1 - \xi_i \end{cases} \tag{13}$$

where $\xi_i \geq 0, \ \ i = 1, 2, ..., n$. The above problem is also described as the dual optimization problem, and the discriminated function can be obtained:

$$d(x) = \mathrm{sgn}(\sum_{x_i \in s_V}^{n} \alpha_i^* y_i K(x_i, x_j) + b^*) \tag{14}$$

Where, $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ is the kernel function,$0 \leq \xi_i \leq C$, $\sum_{i=1}^{n} \alpha_i y_i = 0$ and S_V is the support vector.

3. improved support vector machine.

3.1. improved kernel function. The choice of the kernel function plays an important role in the effect of the classification. It maps the nonlinear data in a low-dimensional space into a high dimension space that is separable. It also does not increase the computation complexity and running time. It uses the kernel function to replace the inner product in the low dimension. The typical kernel function is as follows

(1)The Gaussian kernel function

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

(2)The polynomial kernel function

$$K(x_i, x_j) = ((x_i \cdot x_j) + 1)^d \tag{16}$$

(3)The sigmoid kernel function

$$K(x_i, x_j) = \tanh(v \cdot (x_i \cdot x_j + r)) \tag{17}$$

Where p,d,v, r are real constants, and is the order of the polynomial kernel function. In the actual problem, we choose the specific kernel function and its parameter according to the actual situation.

The composition of the kernel function is made of the local kernel function and global kernel function. The local function is that the data close to the test point, and it can affect the value of the function. Its learning ability is strong and generalization ability is weak. The global kernel function is that the data far from the test point,and it can affect the value of the function. On the contrary, the generalization ability of the global kernel function is strong and its learning ability is weak.

Figure 1 is the curve of the radial basis kernel function when the parameter respectively equals 0.1, 0.2, 0.3, 0.4, and 0.5, while 0.2 is the test point. As we can see from Figure 1, when the input data is near the test point, the value of the kernel changed significantly, indicating that the Gaussian kernel function is the local kernel function.

Figure 2 is the curve of the polynomial kernel function when the parameter d respectively equals 1, 2, 3, 4, 5, while 0.2 is the test point. As we can see from Figure 2, when the input data is far from the test point, the value of the kernel changed significantly, indicating that the polynomial kernel function is the global kernel function.

The polynomial kernel function and the radial basis function are typical representations of the global kernel function and local kernel function, and they have their own limitations. According to the constitute conditions of the kernel function, the sum of the two kernel functions is still a kernel function. Therefore, in this paper, we propose that the polynomial kernel function and the radial basis function combine to form a new kernel function. The learning ability and the generalization ability of the new kernel function are improved significantly. The formula of the new hybrid kernel function is as follows

$$K(x_i, x_j) = \lambda \exp(-\|x_i - x_j\|^2 / p^2) + (1 - \lambda)((x_i \cdot x_j) + 1)^d \tag{18}$$

Where λ is the proportionality coefficient of the two kernel functions in the new hybrid kernel, and the coefficient of the kernel function is between (0,1).

The kernel parameter γ=1/p^2 in the RBF kernel function, the penalty parameter C and the weight coefficient λ in the hybrid kernel function are set generally by multiple experiments or by experience. In this paper, the particle swarm optimization algorithm is used to optimize these parameters.

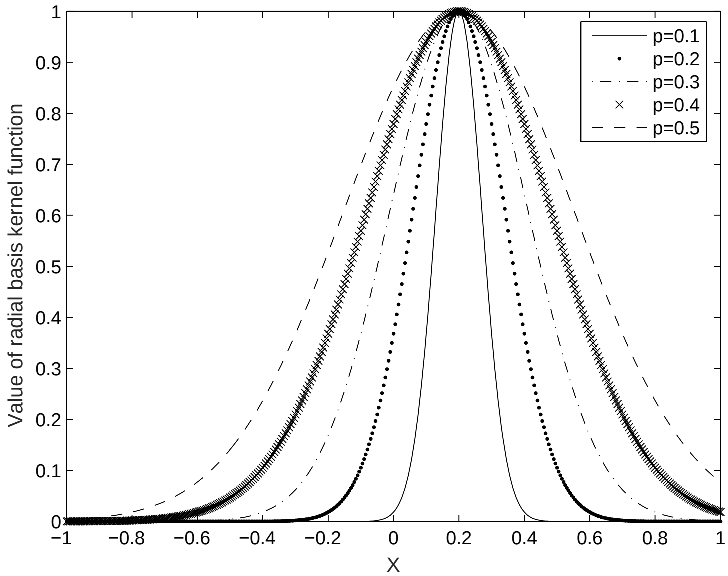

Figure 1. The Curve Of the Radial Basis Kernel Function

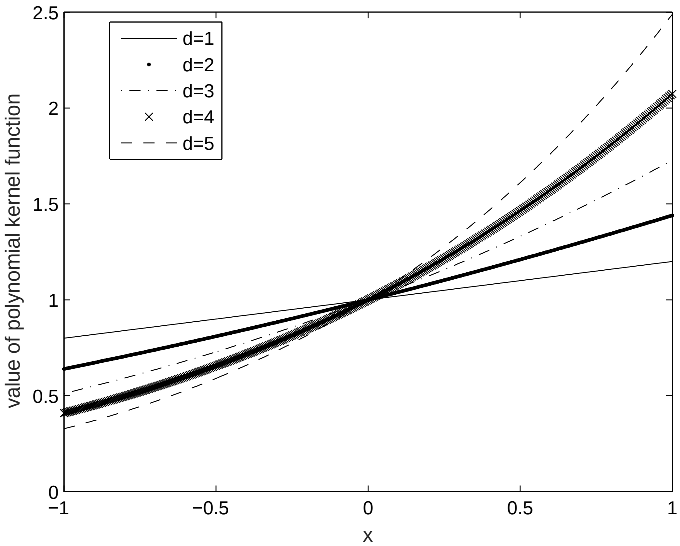

Figure 2. The Curve Of the Polynomial Kernel Function

3.2. **Parameters Optimization.** Particle swarm optimization (PSO) is a population intelligent optimization algorithm. The essence of particle swarm optimization is to determine the optimal position in the problem space by tracking the position of each particle and by collaboration and competition[19]. The parameters that will be optimized are seen as a particle. It flew at a specific speed in the problem space and updated the position and speed of the particle according to Formula (19) and (20). This method can achieve real time tracking of the flight state, and after continuous iterations, it finally finds the position. Suppose that in an n-dimensional space, the population consists of m particles.The position of the first I particle is $x_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$.The velocity of the first I particle is $\mathrm{v}_i = \{v_{i1}, v_{i2}, \ldots, v_{in}\}$.The optimal position of each particle that is

$p_{\rm i} = \{p_{i1}, p_{i2}, \ldots, p_{in}\}$,the global extremum is $p_{\rm g} = \{p_{g1}, p_{g2}, \ldots, p_{gn}\}$.

$$V_i^{t+1} = w * V_i^t + c_1 * rand() * (pbest_i^t - x_i) + c_2 * rand() * (gbest_i^t - x_i) \quad (19)$$

$$x_i^{t+1} = x_i^t + V_i^{t+1} \quad (20)$$

Where w is the inertial multiplier and non-negative.Where w is bigger, the global optimal ability is strong, while the local optimal ability is weak. Where w is smaller, the result is contrary. The $i = 1, 2, ..., M$, M is the quantity of particles in the population. $rand()$ is a number between (0,1), x_i is the current position of the particle, c_1 and c_2 are learning factors. In each dimension, the particle has a maximum limiting speed Vmax. If the velocity is more than the Vmax in a dimension, then the speed in this dimension is restricted to the Vmax (Vmax>0).

When we use the particle swarm algorithm for parameter optimization, set each particle as $x_i = \{C_i, \gamma_i, \lambda_i\}$, and the velocity of the particle is $v_i = \{v_{C_{\rm i}}, v_{\gamma_i}, v_{\lambda_i}\}$. The final optimization goal of the particle is to obtain the highest average recognition accuracy of the disturbance signal. The average recognition accuracy is used as the fitness function in the particle swarm algorithm. The speed and position of the particles are updated by Equation (19) and (20). In the process of iterating, find the optimal parameters c, g, λ. The specifics are as follows:

1) Initialize the particle in the population. Randomly generate the position of each particle $x_i = \{C_i, \gamma_i, \lambda_i\}$, and at the same time, initialize the speed of each particle $v_i = \{v_{C_{\rm i}}, v_{\gamma_i}, v_{\lambda_i}\}$.

2) Calculate the fitness value of the particle. Determine the maximum fitness value, and the value of its corresponding parameter C, r, λ .

3) For each particle, compare its fitness value with the best position Pbest. If its fitness value is better, then we treat the position of the particle as $p_{\rm i}$.

4) For each particle, compare its fitness value with the best position of other particles. If its fitness value is better, then we treat the position of the particle as $p_{\rm g}$.

5) Adjust the velocity and position of the particle according to Equation (19) and (20).

6) If the ending condition is not fulfilled, turn to Step 2. If it reaches the maximum number of iterations, then the calculation is finished and it outputs the result.

4. **Experimental Results and Analysis.** This paper mainly aims at classifying seven kinds of single disturbances such as voltage sag, voltage swell, harmonic, voltage interruption,transient oscillation,transient pulse,flicker and four kinds of multiple disturbances such as voltage sag with harmonic, voltage swell with harmonic,flicker with voltage sag, and flicker with voltage swell. It then validates the effectiveness of the proposed method using MATLAB software. There are 200 samples generated of each disturbance, of which 80 are training samples and 120 are test samples. In order to generate different disturbance signals, the disturbance amplitude, the start time, end time, and duration time are randomly selected according to reference [18]. In the experiment, the fundamental frequency is 50 HZ. There are 1281 points per cycle. The width of signal is 10 cycles, and 15 dB of Gaussian white noise is added.

In the PSO-SVM based on the hybrid kernel function, the value of the individual learning factor is $\rm c_1 = 1.6$, the social learning factor is $\rm c_2 = 1.5$, the range of penalty parameter C is [0.1,100] and the range of nuclear parameter γ is [0.01, 1000]. The maximum of iterations of the particle swarm algorithm is 300 and the population number is 30. The changing process of the fitness value of the particle swarm algorithm is shown in figure 3. Throughout the entire particle swarm optimization process, we use the optimal fitness value of the average classification recognition accuracy of each disturbance as

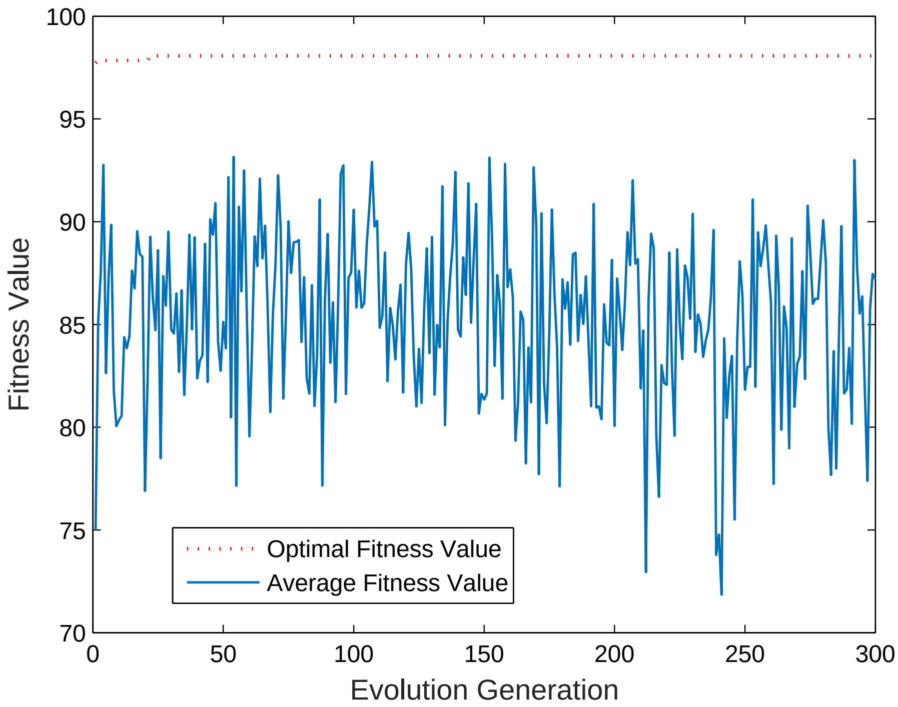

FIGURE 3. The Curve Of the Classification Accuracy Fitness Value

the evaluation criterion. In Figure 3, the blue curve represents the average fitness value of classification accuracy, and the red curve represents the optimal fitness value of the classification accuracy. We can see from Figure 3 that the average fitness value of the classification accuracy varies from 72% to 93%. The optimal fitness value of classification recognition accuracy is 98.1818%, the best fitness value is found on the 25th generation, and the best fitness value is stable after the 25th generation. The final optimal parameter combination is C=2.1755,γ=23.6213,λ=0.5.

Table 1 shows the classification accuracy of the eleven power quality disturbance signals for the SVM classifier based on the radial basis kernel function, the polynomial kernel function, the hybrid kernel function and the PSO-SVM based on hybrid kernel function.It can be seen from Table 1 that for voltage swell, voltage sag, and voltage interruption; the classification accuracy of the proposed method is greatly improved compared with the other three methods. For harmonic and transient oscillation, the classification results of the four methods are all good, reaching 100%. For transient pulse and voltage flicker, the classification accuracy of the proposed method has decreased. For the other four kinds of multiple power quality disturbances, the classification accuracy of the proposed method is greatly improved compared to the other three methods, reaching 100%. For the average classification of the various disturbances signals, the proposed method is 98.1818% higher than the other methods. This means that the proposed method has better performance in classification accuracy than other methods.

In order to more fully compare the difference in performance between the hybrid kernel function and the single kernel function, Table 2 compares the average classification accuracy using the combination of different nuclear parameters, penalty parameters and weight coefficients obtained using the particle swarm algorithm.

It can be seen from Table 2 that the average classification accuracy of the particle swarm optimization based on the hybrid kernel function is higher than that of the single

Table 1. Classification Accuracy of Various Disturbance Signals

Disturbance type	Classification accuracy(%)			
	Radial basis kernel function	Polynomial kernel function	Hybrid kernel function	PSO-HSVM
Voltage swell	99.1667	98.3333	100	100
Voltage sag	70	84.1667	85	99.1667
Voltage interruption	80	72.5	81.6667	99.1667
Harmonic	100	100	100	100
Transient pulse	96.6667	97.5	99.1667	88.3333
Transient oscillation	100	100	100	100
Voltage flicker	99.1667	100	100	93.3333
Voltage swell with harmonic	98.3333	100	100	100
Voltage sag with harmonic	99.1667	97.5	98.3333	100
Voltage swell with flicker	97.5	97.5	99.1667	100
Voltage sag with flicker	98.3333	95.8333	95	100
Average classification accuracy(%)	94.3939	94.8485	96.2121	98.1818

Table 2. Parameters and Classification Accuracy of Different Algorithms

	C	g	λ	Accuracy(%)
POLY kernel function	3.9863			94.8485
RBF kernel function PSO-SVM	1.4142	9		96.3636
RBF kernel function PSO-SVM	1	0.7071		95.9848
RBF kernel function PSO-SVM	5.6569	2.8284		95.7576
Hybrid kernel function PSO-SVM	46.44	25.5406	0.74	98.1061
Hybrid kernel function PSO-SVM	2.8017	35.2934	0.61	97.4242
Hybrid kernel function PSO-SVM	84.2248	0.8948	0.42	97.803
Hybrid kernel function PSO-SVM	2.1755	23.6213	0.5	98.1818

kernel function. The range of parameter C,γ using the radial basis kernel function is small. However, the range of parameter C,γ using the hybrid kernel function is larger, and it can obtain the various groups of high accuracy parameter combinations. For the single kernel function, the range of parameters with higher classification accuracy is relatively fixed. For the hybrid kernel function, it is not confined to a small range and has a multi-group parameter combination with higher classification accuracy. It shows that PSO-SVM based on the hybrid kernel function has better learning ability and generalization ability.

5. **Conclusions.** This paper proposes a method for identifying the multiple power quality disturbance based on PSO-SVM of hybrid kernel function. Wavelet transform is used to extract wavelet energy difference as a feature vector. PSO-SVM based on hybrid kernel

function is then used to classify the wavelet energy difference. The hybrid kernel function maximizes the learning ability of RBF kernel function and the generalization ability of polynomial kernel function by adjusting the weight coefficient. We use the particle swarm to optimize the parameters in the classifier model to find the optimal parameter combination. The simulation result shows that the average classification accuracy of the proposed method is feasible and higher than that of single kernel function.

REFERENCES

[1] Q. G. Zhang, Z. G. Liu, and L. Liu, Recognition of multiple power quality disturbances using multi-label wavelet support vector machine, *Proceedings of the CSEE*, vol.33, no.28, pp.114–120, 2013.

[2] J. B. Li, M. Li, J. S.Pan, S. C.Chu, and J. F.Roddick, Gabor-based kernel self-optimization Fisher discriminant for optical character segmentation from text-image-mixed document, *Optik-International Journal for Light and Electron Optics*, vol.126, no.21, pp.3119–3124, 2015.

[3] X. D.Xie, and B. H. Li, A Unified Framework of Multiple Kernels Learning for Hyperspectral Remote Sensing Big Data, *Journal of Information Hiding and Multimedia Signal Processing*, vol.7, no.2, pp.296–303, 2016.

[4] J. S.Pan, J. B. Li, and Z. M.Lu, Adaptive Quasiconformal Kernel Discriminant Analysis, *Neurocomputing*, vol.71, no.13-15, pp.2754–2760, 2008.

[5] J. B. Li, J. S.Pan, and Z. M.Lu, Kernel Optimization-Based Discriminant Analysis for Face Recognition, *Neural Computing and Applications*, vol.18, no.6, pp.603–612, 2009.

[6] J. S.Pan, J. B. Li, and Z. M.Lu, Overview of Deep Kernel Learning Based Techniques and Applications, *Journal of Network Intelligence*, vol.1, no.3, pp.82–97, 2016.

[7] Y. H.Gu, and M. H.Bollen, Time-frequency and time-scale domain analysis of voltage disturbances, *IEEE Transactions on Power Delivery*, vol.15, no.4, pp. 1279–1284, 2010.

[8] S. Khokhar, A. M.Zin, and A. P.Memon, A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network, *Measurement*, vol.95, pp.246–259, 2017.

[9] D.D. 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, pp.6075–6081, 2015.

[10] V.Laparra, U. Michael, G. J. Malo , and A. Hyvarinen, Power Quality Disturbance Classification Using the S-Transform and Probabilistic Neural Network, *Energies*, vol 10, pp.1–19, 2017.

[11] H. M.Faridah, M. A.Rahman, M. F.Yaakub, Performance of Modified S-Transform for Power Quality Disturbance Detection and Classification, *Telkomnika*, vol.15, no.4, pp.1520–1529, 2017.

[12] G. Han, J. W.Zhang, and X.Chun, 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.

[13] Y. J. Wang, Y. Li, and Z. W. Qu, The identification of power quality disturbance based on PSO MP algorithm and RBF neural network, *Electrical Measurement & Instrumentation* , vol.53, no.13, pp.54–58, 2016.

[14] H. F. Chen, and G. Y. Zhang Power quality disturbance using decision tree and support vector machine, *Power System Technology*, vol.37, no.5, pp.69–72, 2013.

[15] M. Y. Wang, X.J.Wang, and T.Y.Li, Research on classification of power quality based on norm entropy of morphological wavelet and support vector machine, *Electric Machine and Control*, vol.20, no.9, pp.33–39, 2016.

[16] T. X. Cui, X. L.Zhou, and W. H.Liu, Gear Fault Diagnosis based on Hilbert Envelope Spectrum and SVM, *Journal of Northeast Electric Power University*, vol.37, no.6, pp.56–61, 2017.

[17] M. Yang, X. X.Chen, Q. Zhang, D. Y. Li, Y.Sun, and Y. P.Jia, A Review of short-term wind speed prediction based on support vector machine, *Journal of Northeast Electric Power University*, vol.37, no.4, pp.1–7, 2017.

[18] L. Q. Zhao, and Y. Long, Multiple power quality disturbances classification based on the improved SVM, *Advanced Technology of Electrical Engineering and Energy*, vol.35, no.7, pp.44–49, 2016.

[19] S. B.Wang, J. L.Zhou, J. S Wang, and T. T.Deng, Particle Swarm Algorithm and Its Application in Parameter Identification of Steam Turbine Governing System, *Journal of Northeast Electric Power University*, vol.37, no.5, pp.51–55, 2017.