

Classification of Power Quality Disturbances Using Wavelet Packet Energy Entropy and LS-SVM

Ming Zhang, Kaicheng Li, Yisheng Hu

College of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China

E-mail: zmcoc@yaho.com.cn

Received April 11, 2010; revised May 22, 2010; accepted June 27, 2010

Abstract

The power quality (PQ) signals are traditionally analyzed in the time-domain by skilled engineers. However, PQ disturbances may not always be obvious in the original time-domain signal. Fourier analysis transforms signals into frequency domain, but has the disadvantage that time characteristics will become unobvious. Wavelet analysis, which provides both time and frequency information, can overcome this limitation. In this paper, there were two stages in analyzing PQ signals: feature extraction and disturbances classification. To extract features from PQ signals, wavelet packet transform (WPT) was first applied and feature vectors were constructed from wavelet packet log-energy entropy of different nodes. Least square support vector machines (LS-SVM) was applied to these feature vectors to classify PQ disturbances. Simulation results show that the proposed method possesses high recognition rate, so it is suitable to the monitoring and classifying system for PQ disturbances.

Keywords: Power Quality (PQ), Wavelet Packet Transform (WPT), Wavelet Packet Log-Energy Entropy, Least Square Support Vector Machines (LS-SVM)

1. Introduction

The deregulation policies in electric power systems results in the absolute necessity to quantify power quality (PQ). This fact highlights the need for an effective recognition technique capable of detecting and classifying the PQ disturbances. Traditionally PQ recordings are analyzed in the time-domain by skilled engineers. However, PQ disturbances may not always be obvious in the original time-domain signal. One of the traditional signal processing techniques called Fourier transform provides information in frequency-domain but it does have limitations. One crucial limitation is that a Fourier coefficient represents a component that lasts for all time. This makes Fourier analysis less suitable for non-stationary signals. Wavelet analysis, which provides both time and frequency information, can overcome this limitation. Unlike the Fourier transforms, the wavelet transform has a fully scalable window, which allows a more accurate local description and separation of signal characteristics [1]. The wavelet transform has been applied to the wide range of PQ signals analysis: feature extraction [2], noise reduction [3], and data compression [4]. Recently, The identification of PQ disturbances is often based on arti-

cial neural network (ANN) [5], fuzzy method (FL) [6], expert system (ES) [7], support vector machines (SVM) [8], and hidden Markov model (HMM) [9]. Many of the studies proposed in the literature present that these techniques can use feature vectors derived from disturbance waveforms to classify PQ disturbances.

The types of PQ disturbances include the sag, interruption, swell, harmonic, notch, oscillatory transient (Osc. transient) and impulsive transient (Imp. transient) (see **Figure 1**) [10]. In this paper, the combined technique of wavelet packet transform (WPT) and least square support vector machines (LS-SVM) for PQ disturbances recognition is presented. Decision making is performed in two stages: feature extraction and LS-SVM as a classifier. **Figure 2** shows the block diagram of the classification system. The details of each stage are described in the next sections. High accuracies were achieved by using the LS-SVM trained on the wavelet packet log-energy entropy of different nodes.

The rest of this paper is organized as follows. In Section 2, the feature extraction by WPT is explained. In Section 3, brief review of the LS-SVM with the minimum output coding (MOC) technique is presented.

In Section 4, the results of classification of the LS-SVM trained on wavelet packet log-energy entropy to

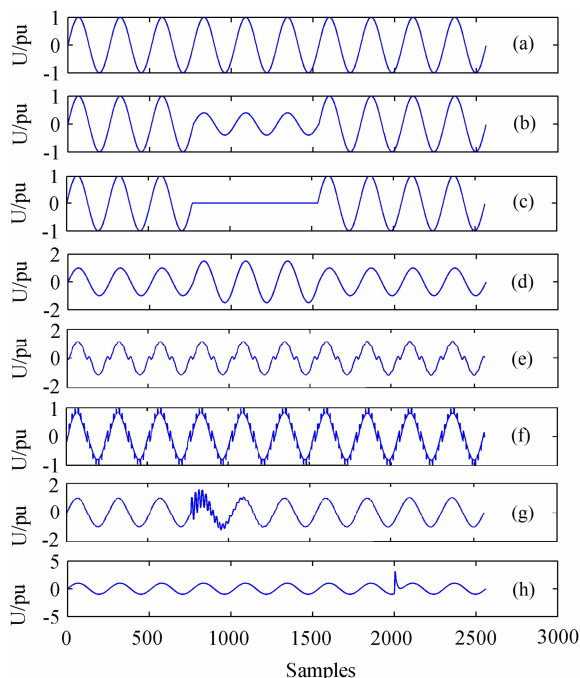


Figure 1. Power quality disturbance waveforms: (a) Normal signal; (b) Sag; (c) Interruption; (d) Swell; (e) Harmonic; (f) Notch; (g) Oscillatory transient; (h) Impulsive transient.

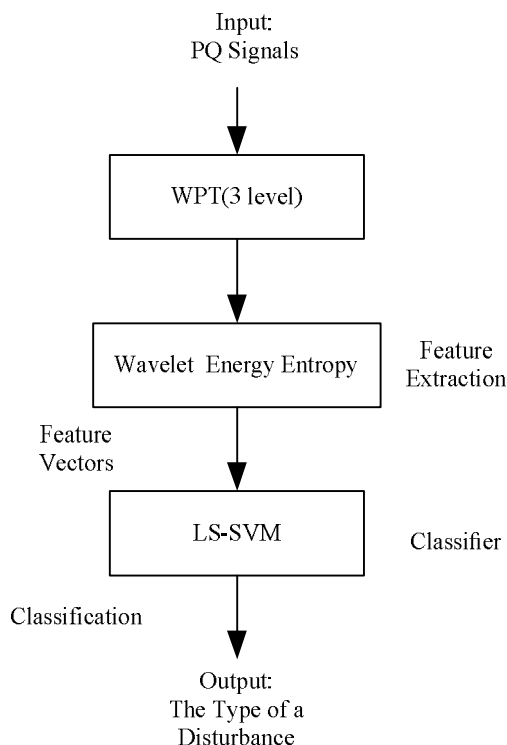


Figure 2. Block diagram of the classification system.

the studied PQ disturbance signals are presented. Finally, conclusions are given in Section 5.

2. Feature Extraction Using WPT

The purpose of the feature extraction process is to select and retain relevant information from original signals. The WPT was first applied to decompose the original PQ signals into frequency bands. One of the advantages of the WPT is that it is able to decompose signals at various resolutions, which allows accurate feature extraction from non-stationary signals like PQ disturbances. The features of signals, such as wavelet packet energy entropy, were then extracted from these decomposed signals as feature vectors.

The wavelet transform decomposes a signal into a set of basic functions called wavelets. These basic functions are obtained by dilations, contractions and shifts of a unique function called wavelet prototype. Continuous wavelets are functions generated from one single function by dilations and translations of a unique admissible mother wavelet $y(t)$:

$$y_{a,b}(t) = \frac{1}{\sqrt{|a|}} y\left(\frac{t-b}{a}\right) \quad (1)$$

where $a, b \in \mathfrak{R}, a \neq 0$ are the scale and translation parameters, respectively, and t is the time. The function set $(y_{a,b}(t))$ is called wavelet family. It is common to employ both wavelet and scaling functions in the transform representation. In general, the scale and shift parameters of the discrete wavelet family are given by $a = a_0^j$ and $b = kb_0 a_0^j$, where j and k are integers. The function family with discretized parameters becomes:

$$y_{j,k}(t) = a_0^{-j/2} y(a^{-j}t - kb_0) \quad (2)$$

where $y_{j,k}(t)$ is called the discrete wavelet transform (DWT) basis.

DWT analyzes the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions $j(t)$ and wavelet functions $y(t)$, which associated with low-pass and high-pass filters, respectively. The original signal $x(t)$ can be decomposed to:

$$x(t) = \sum_k c_j(k) j_{jk}(t) + \sum_{j=1}^J \sum_k d_j(k) y_{jk}(t) \quad (3)$$

where j is the level number of the wavelet decomposition, $j = 1, 2, \dots, J$ with J the time of the wavelet decomposition. c_j and d_j are the approximation coefficients and detail coefficients of $x(t)$, respectively.

Because the information in higher frequency components is important, the frequency resolution of DWT may

not be fine enough to extract pertinent frequency information about the signal. The necessary frequency resolution may be achieved by using WPT, an extension of the DWT. In the WPT, the wavelet detail at each level is, in addition to decomposition of only the wavelet approximation in the regular wavelet analysis, further decomposed in to its own approximation and detail components. By this process, some lower frequency contents leaked in the wavelet details at the previous level can be further sifted out at the current level and also the frequency resolution for signal analysis increases. As a result, the WPT may provide better accuracy in both higher and lower frequency components of the signal.

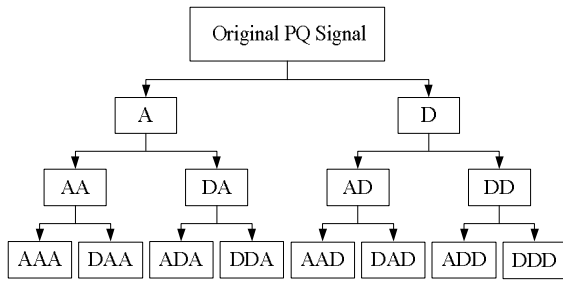
Figure 3 shows the wavelet packet decomposition tree for three levels ($J = 3$). For each level of decomposition the signal is filtered into approximate information of the signals (lower frequency component) and detail information (higher frequency component). If this procedure is repeated J times, a filter bank is created with J filters.

To evaluate the importance of the wavelet packet components to a signal, the concept of entropy is often applied in signal processing and there are various definitions of entropy in the literature. Among them, two representative ones are used in the present article, *i.e.* the energy entropy and the Shannon entropy. The wavelet packet energy entropy at a particular node n in the wavelet packet tree of a signal is a special case of $p = 2$ of the p -norm entropy, defined as

$$Ent_n = \sum_k |wc_{n,k}|^p \quad (p \geq 1) \quad (4)$$

where $wc_{n,k}$ denotes the wavelet packet coefficients corresponding to node n at time k . It was demonstrated that the wavelet packet energy has more potential for use in signal classification as compared to the wavelet packet coefficients alone. The wavelet packet energy represents energy stored in a particular frequency band and is mainly used in this study to extract the dominant frequency components of the signal.

The Shannon energy entropy and relative Shannon energy entropy are defined respectively as [11]



A: Approximate info of Signals (Low Frequency component)

D: Detail info of Signals (High Frequency component)

Figure 3. Wavelet packet decomposition tree.

$$Ents_n = -\sum_k wc_{n,k}^2 \log(wc_{n,k}^2) \quad (5)$$

$$REnts_n = Ents_n / Ents_{nor_n} \quad (6)$$

where $Ents_{nor_n}$ is the Shannon energy entropy of the normal signal corresponding to node n .

In this paper, one of the commonly used entropy, log-energy entropy is also defined as

$$Entl_n = \sum_k \log(wc_{n,k}^2) \quad (7)$$

The relative log-energy entropy is proposed as

$$REntl_n = Entl_n / Entl_{nor_n} \quad (8)$$

where $Entl_{nor_n}$ is the log-energy entropy of the normal signal corresponding to node n .

3. LS-SVM

The second stage is the disturbances classification. Support vector machine (SVM) can avoid the problems of over learning, dimension disaster and local minimum in the classical study method, and is applied in many classification problems successfully [8,11]. According to the practice, [12] advanced by J. A. K. Suyken can overcome the disadvantage of slow training velocity in the large scale problem, as LS-SVM algorithm translates the quadratic optimization problem into that of solving linear equation set. Although a wide range of classifiers are available, we use LS-SVM in this paper.

We consider a training set of N data points $\{x_k, y_k\}$, $k = 1, 2, \dots, N$, where $x_k \in \mathfrak{R}^n$ is the input data, $y_k \in \mathfrak{R}$ is the k -th output data, the SVM constructs a decision function that is represented by:

$$y(x) = w^T x + b \quad (9)$$

where the dimension of w is not specified. It means that it can be infinitely dimensional. The separating hyperplane that creates the maximum distance between the plane and the nearest data is called as the optimal separating hyperplane as shown in **Figure 4**.

In LS-SVM for the function estimation the following optimization problem can be given

$$\min_{w,b,e} J_{LS}(w, b, e) = \frac{1}{2} w^T w + C \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (10)$$

subject to the equality constraints

$$y_k = w^T x_k + b + e_k, \quad k = 1, \dots, N \quad (11)$$

where e_k are slack variables and C is a positive real constant. One defines the Lagrangian

$$L(w, b, e; a) = J_{LS} - \sum_{k=1}^N a_k (w^T x_k + b + e_k - y_k) \quad (12)$$

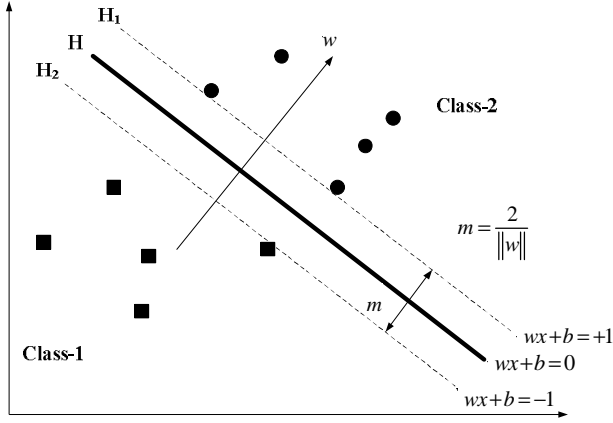


Figure 4. Optimal separating hyper plane.

with Lagrange multipliers a_k . The conditions for optimality are

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N a_k x_k \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N a_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow a_k = g e_k \\ \frac{\partial L}{\partial a_k} = 0 \rightarrow w^T x_k + b + e_k - y_k = 0 \end{cases} \quad (13)$$

for $k = 1, 2, \dots, N$. It can be written immediately as the solution to the following set of linear equations:

$$\begin{bmatrix} \mathbf{I} & 0 & 0 & -\mathbf{X} \\ \mathbf{I} & 0 & 0 & -\mathbf{1}^T \\ \mathbf{I} & 0 & g\mathbf{I} & -\mathbf{I} \\ \mathbf{X} & \mathbf{1} & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ \mathbf{e} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{Y} \end{bmatrix} \quad (14)$$

with $\mathbf{X} = [x_1, \dots, x_N]$, $\mathbf{Y} = [y_1, \dots, y_N]$, $\mathbf{1} = [1, \dots, 1]$, $\mathbf{e} = [e_1, \dots, e_N]$ and $\mathbf{a} = [a_1, \dots, a_N]$. The solution is finally given by

$$\begin{bmatrix} 0 & -\mathbf{1}^T \\ \mathbf{1} & \mathbf{X}^T \mathbf{X} + g^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix} \quad (15)$$

with $w = \sum_k a_k x_k$, $e_k = a_k / C$. The support values a_k are proportional now to the errors at the data points.

So far we explained the linear case. SVM's with polynomials, splines, radial basis function networks, or multilayer perceptrons as kernels are obtained after mapping the input data into a higher dimensional space by $f(x_k)$, where $f(\cdot) : \mathcal{R}^n \rightarrow \mathcal{R}^{n_h}$. The number n_h does not have to be specified because of the application of

Mercer's condition, which means that

$$K(x_k, x_j) = f(x_k)^T f(x_j) \quad (16)$$

can be imposed for these kernels. Finally, the nonlinear function takes the form:

$$y(x) = \sum_{k=1}^N a_k K(x, x_k) + b \quad (17)$$

where the parameters a_k , b follow from (15) after replacing $x_k^T x_j$ by $K(x_k, x_j)$.

Multi-class classification was realized by the combination of LS-SVM classifiers with the minimum output coding (MOC) technique. In the MOC technique, up to \log_2^m (where m is the number of classes) LS-SVM classifiers were trained, and each of them aimed to separate a different combination of classes. There were eight classes (normal signal, sag, interruption, swell, harmonic, notch, oscillatory transient and impulsive transient) in this study, so three classifiers were necessary to differentiate them. The coding was defined by the codebook represented by a matrix, where the columns represent the different classes, and the rows indicate the results of the binary classifiers. The multi-class classifier output code for a pattern is a combination of targets of these three classifiers. In this study, the eight classes were encoded in the following codebook of minimum output coding:

	C1	C2	C3	C4	C5	C6	C7	C8
codebook =	-1	-1	-1	-1	1	1	1	1
	-1	-1	1	1	-1	-1	1	1
	-1	1	-1	1	-1	1	-1	1

where C1, C2, C3, C4, C5, C6, C7, and C8 are normal signal, sag, interruption, swell, harmonic, notch, oscillatory transient and impulsive transient, respectively.

4. Simulation Analysis

To test classification results for PQ disturbances, the testing samples of these PQ disturbances have been generated using algebraic equations [14]. The advantage of using algebraic equations for evaluation is the flexibility of adjusting signal noise contents as well as various waveform parameters such as the disturbance occurrence time, harmonic contents, sag depth, etc.

These disturbance waveforms are generated at a sampling rate of 256 samples/cycle for a total of 2560 points (10 cycles). In order to create different disturbance cases, some unique parameters such as starting time, magnitude, duration, frequency, and damping are allowed to change randomly. The random generation of signals is helpful for the testing of the classification more reliable since none of these attributes is fixed for real distribution sys-

tem disturbances.

Using wavelet packet decomposition, each signal shown above was decomposed to level 3. The wavelet ‘Daub4’ was selected because it is more adequate for classification of PQ disturbances [13]. The wavelet packet energy entropy of different nodes of the decomposed signals were calculated, which could be used to identify the type of PQ disturbances. The performances of difference wavelet packet energy entropy for feature sets are shown in **Figure 5**. From above **Figure 5**, we can conclude that relative log-energy entropy is more effective than traditional relative Shannon energy entropy, which can amplify the errors among the feature vectors. These features consist of 8-dimension feature space.

In this paper, we construct a LS-SVM by using radial basis function (RBF) as kernel function in LS-SVM pro-

posed above.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2s^2}\right) \quad (18)$$

where s is the width of the kernel.

For training the SVMs with RBF kernel functions, one has to predetermine the s values. The optimal or near optimal s values can only be ascertained after trying out several, or even many values. Beside this, the choice of C parameter in the SVM is very critical in order to have a properly trained SVM. The SVM has to be trained for different C values until to have the best result. From the **Figure 6**, It is found that the near optimal values are $s^2 = 1$ and $C = 4$.

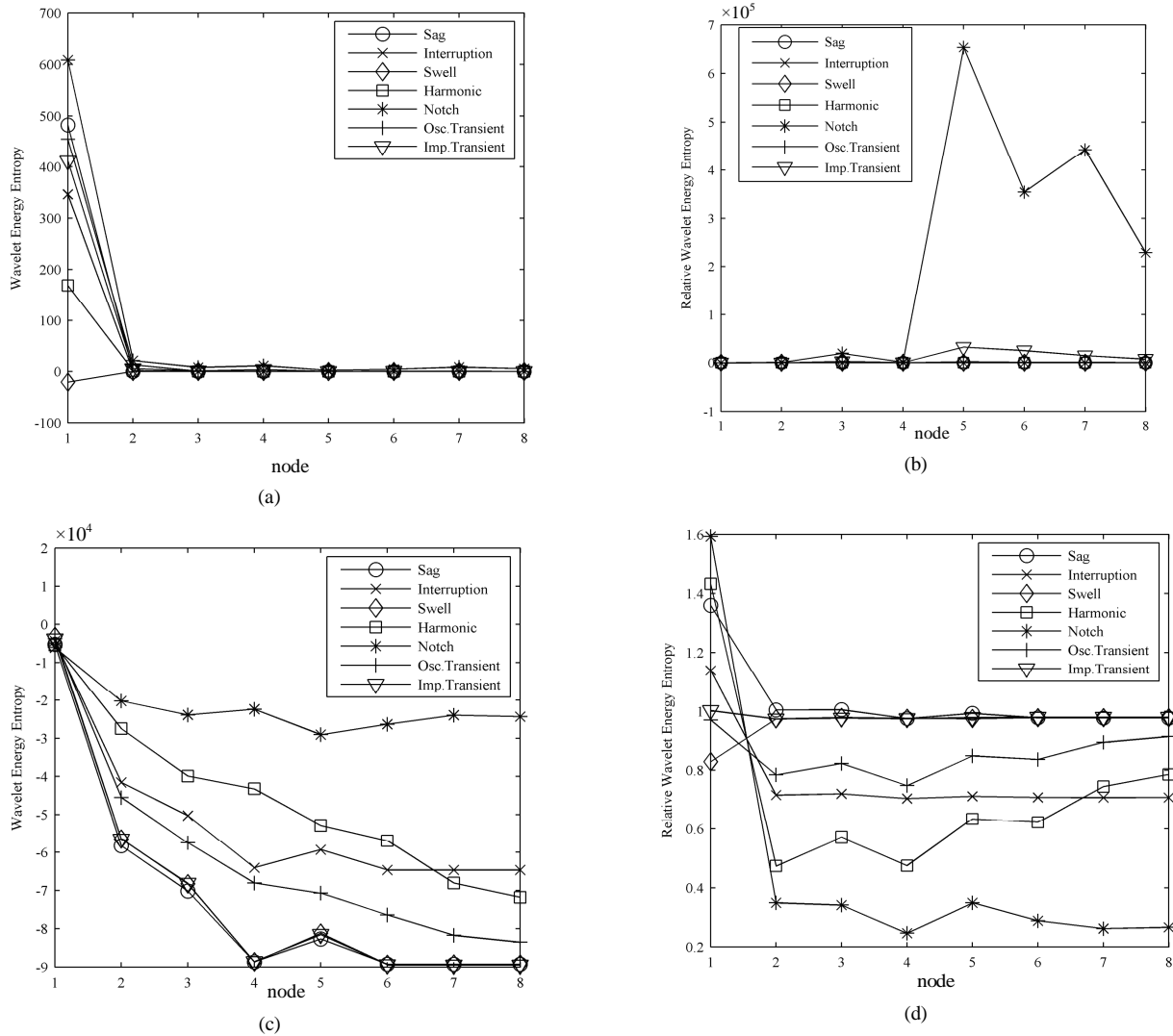


Figure 5. Performance comparison of difference wavelet energy entropy of the waveforms in Figure 1: (a) Wavelet packet Shannon energy entropy; (b) Relative wavelet packet Shannon energy entropy; (c) Wavelet packet log-energy entropy; (d) Relative wavelet packet log-energy entropy.

Each decomposed signal now has eight features ($J = 3$). The feature vectors of PQ disturbances are fed to the LS-SVM for classification. The LS-SVM topology used for classification is shown in **Figure 7**. We trained three different LS-SVMs (LS-SVM1, LS-SVM2, LSSVM3) for seven different PQ disturbances (seven hundred samples of various PQ disturbances). The patterns to be distinguished from others are represented by +1 and the remaining patterns represented by -1 for both training and testing procedures.

The output of three different LS-SVMs constructs the code of the input PQ signals, which the type of a disturbance or the normal signal will be identified. In the present work a standard feed-forward network with 8 input neurons, 12 hidden neurons, and 7 output neurons was compared to the LS-SVM implementation. Furthermore, our results indicate that solutions obtained by LS-SVM training seem to be more robust with a smaller standard

error compared to standard ANN training using the same features as inputs.

The other seven hundred PQ disturbances of various types have been generated for the testing. The classification results in a correct identification rate of 97.7% are shown in **Table 1** using the proposed LS-SVM classifier. For comparison purposes, the total classification accuracies on the same test sets and the CPU times of training of the two classifiers are presented in **Table 2**. It is found that the proposed LS-SVM classifier performed better than the standard ANN classifier.

To evaluate the performance of the kernel function, three LS-SVM classifiers were developed based on the linear kernel, the polynomial kernel, and the RBF kernel. The classification results with linear, polynomial and RBF kernel are shown in **Table 3**. The accuracy of classification is high in RBF kernel in comparison with the polynomial and linear kernels.

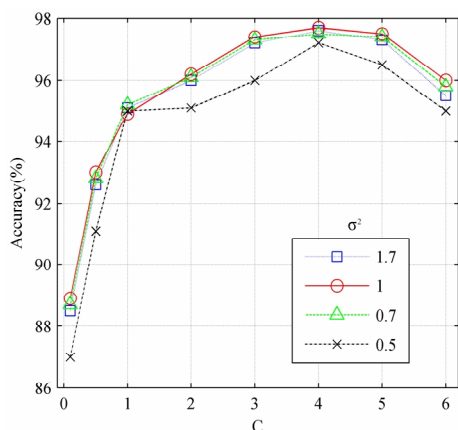


Figure 6. Comparison of accuracy acquired with different C and σ^2 values for RBF kernels.

Table 2. Comparison of the classification indices between the LS-SVM and ANN classifiers.

Classifier	Training set samples	Testing set samples	Mean training time (s)	Mean testing time (s)	Mean correct ratios (%)
LS-SVM	700	700	9.968	1.922	97.7
ANN	700	700	101.523	1.993	95.2

Table 1. Classification results using the proposed LS-SVM classifier.

Type of PQ disturbances	Number of disturbances	Number of disturbances classified	Number of disturbances misclassified	Classification Accuracy (%)
Sag	100	97	3	97
Interruption	100	97	3	97
Swell	100	99	1	99
Harmonic	100	98	2	98
Notch	100	99	1	99
Osc. transient	100	97	3	97
Imp. transient	100	96	4	96
Sum	700	684	16	97.7

Table 3. Classification accuracies for the different kernels used.

Kernel used	Number of disturbances in training	Number of disturbances in testing	Number of disturbances misclassified	Classification accuracy (%)
Linear	700	700	27	96.1
Polynomial	700	700	20	97.1
RBF	700	700	16	97.7

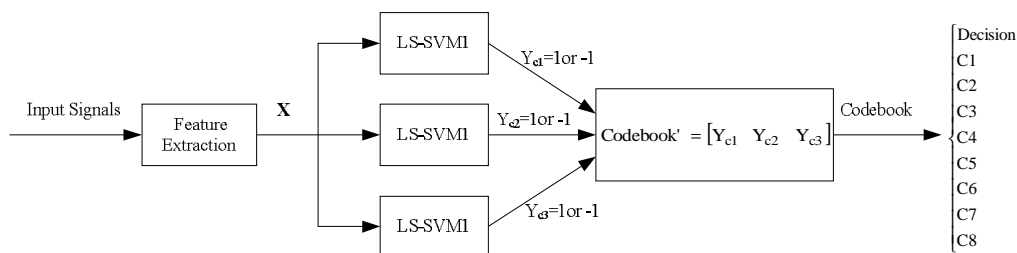


Figure 7. Classification of PQ disturbances based on MOC (Codebook' is one column of Codebook).

5. Conclusions

In this paper, an attempt has been made to extract efficient features of the PQ disturbances using WPT and to classify the disturbances using LS-SVM with the MOC technique. It is also found that relative wavelet packet log-energy entropy is considered as feature vectors, which are suitable for classification of PQ disturbances. For comparison different classifiers, the LS-SVM and ANN classifiers were implemented to deal with the same classification. The classification accuracies and the CPU times of training showed that the LS-SVM classifier produces considerably better performance than that of the ANN classifier.

6. Acknowledgements

The authors would like to thank to the support of Wuhan Xinlian Science and Technology Ltd.

7. References

- [1] S. Mallat, "A Wavelet Tour of Signal Processing," Academic Press, San Diego, California, 1998.
- [2] S. Santoso, E. J. Powers and P. Hofman, "Power Quality Assessment via Wavelet Transform Analysis," *IEEE Transaction on Power Delivery*, Vol. 11, No. 2, 1996, pp. 924-930.
- [3] H. T. Yang and C. C. Liao, "A De-Noising Scheme for Enhancing Wavelet-Based Power Quality Monitoring System," *IEEE Transaction on Power Delivery*, Vol. 16, No. 3, 2001, pp. 353-360.
- [4] S. Santoso, E. J. Powers and W. M. Grady, "Power Quality Disturbance Data Compression Using Wavelet Transform Methods," *IEEE Transaction on Power Delivery*, Vol. 12, No. 3, 1997, pp. 1250-1257.
- [5] A. K. Ghosh and D. L. Lubkeman, "The Classification of Power System Disturbance Waveforms Using a Neural Network Approach," *IEEE Transaction on Power Delivery*, Vol. 10, No. 1, 1995, pp. 109-115.
- [6] T. X. Zhu, S. K. Tso and K. L. Lo, "Wavelet-Based Fuzzy Reasoning Approach to Power Quality Disturbance Recognition," *IEEE Transaction on Power Delivery*, Vol. 19, No. 4, 2004, pp. 1928-1935.
- [7] M. B. I. Reaz, F. Choong, M. S. Sulaiman, F. Mohd-Yasin and M. Kamada, "Expert System for Power Quality Disturbance Classifier," *IEEE Transaction on Power Delivery*, Vol. 22, No. 3, 2007, pp. 1979-1988.
- [8] P. Janik and T. Lobos, "Automated Classification of Power Quality Disturbances Using SVM and RBF Networks," *IEEE Transaction on Power Delivery*, Vol. 21, No. 3, 2006, pp. 1663-1669.
- [9] J. Chung, E. J. Powers, W. M. Grady and S. C. Bhatt, "Power Disturbance Classifier Using a Rule-Based Method and Wavelet Packet-Based Hidden Markov Model," *IEEE Transaction on Power Delivery*, Vol. 17, No. 1, 2002, pp. 233-241.
- [10] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE Standards Description: 1159-1995, 2009.
- [11] G. S. Hu, F. F. Zhu and Z. Ren, "Power Quality Disturbance Identification Using Wavelet Packet Energy Entropy and Weighted Support Vector Machines," *Expert Systems with Applications*, Vol. 35, No. 1-2, 2008, pp. 143-149.
- [12] J. A. K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Processing Letter*, Vol. 9, No. 3, 1999, pp. 293-300.
- [13] N. S. D. Brito, B. A. Souza and F. A. C. Pires, "Daubechies Wavelets in Quality of Electrical Power," *8th International Conference on Harmonics and Quality of Power*, Athens, 14-18 October 1998, pp. 511-515.
- [14] T. K. Abdel-Galil, M. Kamel, A. M. Youssef, E. F. El-Saadany and M. M. A. Salama, "Power Quality Disturbance Classification Using the Inductive Inference Approach," *IEEE Transaction on Power Delivery*, Vol. 19, No. 4, 2004, pp. 1812-1818.