Power quality event characterization using support vector machine and optimization using advanced immune algorithm

B. Biswal a,*, M.K. Biswal b, P.K. Dash c, S. Mishra d

a GMR Institute of Technology, Rajam, Srikakulam, AP 532127, India
b Silicon Institute of Technology, Bhubaneswar, India
c S’O’A University, Bhubaneswar, India
d Indian Institute of Technology, Delhi, India

A R T I C L E   I N F O

Article history:
Received 24 October 2011
Received in revised form 21 March 2012
Accepted 6 August 2012
Communicated by A. Zobaa
Available online 12 October 2012

Keywords:
Power signals
TT-transform
Support vector machine (SVM)
Radial basis function
Mexican hat wavelet kernel
Modified immune optimization algorithm (MIOA)

A B S T R A C T

This paper presents a time-time transform (TT-transform) variant for classification of non-stationary power signal disturbance patterns. The TT-transform variant is derived from the well known S-transform and employs a new window function whose width is inversely proportional to the frequency raised to a constant power with values within 0 and 1. Features are derived from the TT-transform result of the power signal patterns. These features are used for automatic recognition of types of disturbances with the help of kernel based support vector machine (SVM) based clustering. Further, the clustering performance of the TT-SVM based pattern recognizer is improved by a modified immune optimization algorithm. Several test cases are provided to demonstrate the improvement in classification accuracy while resulting in significant reduction of support vectors.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The voltage and current signals, in electrical power networks, often exhibit fluctuations in amplitude, phase, and frequency due to the operation of solid-state devices that are prolifically used for power control. This gives rise to steady state as well as transient voltage disturbances like voltage sag, voltage swell, interruption, oscillatory transients, impulsive transients, multiple voltage notches due to solid-state converter switching and harmonics. Power sinu-
soids being modulated by low frequency signals are also observed in the electric power networks. To distinguish between these disturbance signal patterns in the normal sinusoidal signals of frequency 50 Hz or 60 Hz, advanced signal processing techniques along with intelligent system approach play an important role.

The detection and localization of these disturbances are critical in order to determine the causes and sources of disturbances. Harmonics, which are major phenomena in power networks are well detected using neural networks and adaptive filters [1,2]. In the current research trends wavelet transform (WT) [3–7] is widely used in analyzing non-stationary signals for power quality assessment. Although WT has been extensively used for detection of power quality disturbances, the effect of electrical noise is not adequately considered in many of the cases. In [5] it is shown that when noise is present in the signal wavelet based MLP neural network performs badly in comparison with other time-frequency techniques for power quality event classification. A de-noising scheme has been proposed in [6] for enhancing wavelet based power quality monitoring system. Apart from WT, the short time Fourier transform (STFT) and S-transform (ST) are often preferred for Power signal classification.

The main objective of this paper is to apply TT-transform (TT) [8] with a modified Gaussian window and SVM clustering to the problem of power signal classification. The TT-transform is derived from the ST by taking the inverse Fourier transform along the row vectors of ST result, corresponding to frequency axis. Thus the TT transform results in a time-time domain decomposition of the signal pattern. The first paper in the area of power quality analysis using S-transform [9] uses a Gaussian window with one scaling factor. In the present paper the shape of the Gaussian window is slightly modified to have a better time and frequency resolution. Correspondingly the TT transform with the modified Gaussian window provides an improved time–frequency resolution. Further, the invertibility of the TT-transform leads to the possibility of filtering and signal to noise improvements in the time domain.

After the features are extracted from the time-series data using modified TT-transform a SVM algorithm with the choice of different kernel functions, is used to group the data into clusters and thereby identifying belongingness of the data to a particular class. The SVM
The S-transform often suffers from poor concentration of energy at higher frequency and hence poor frequency localization. So the dependence of standard deviation of the Gaussian window on frequency can be altered to improve energy concentration.

Defining the standard deviation of the window as

$$
\sigma(f) = \frac{k}{|f|} \quad k > 0 \quad \text{and} \quad 0 \leq c \leq 1
$$

Resulting in a modified S-transform as

$$
S(t,f) = \frac{1}{k \sqrt{2\pi}} \int_{-\infty}^{\infty} u(t) e^{-\frac{(t-t')^2}{2\sigma^2}} e^{-2\pi j ft} dt
$$

The parameter 'c' is initially introduced as a switch between STFT (c=0) and S-transform (c=1) but later it was observed that better localization of the time-frequency distribution can be obtained by varying 'c' value.

Here, c < 0 cannot be taken, since on doing so σ instead of being proportional to the time period of the signal becomes inversely proportional to it, which takes it outside the basic definition of S-transform.

We define a modified time–time transform or TT-transform by taking the inverse Fourier transform of the modified S-transform along the frequency vectors.

$$
TT(t,t) = \int_{-\infty}^{\infty} S(t,f) e^{2\pi j ft} df
$$

Also inverting the TT transform, the original signal \( u(t) \) is obtained as

$$
u(t) = \int_{-\infty}^{\infty} TT(t,t) dt
$$

In the discrete domain the modified TT transform is expressed as

$$
\hat{TT}(j,k) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}} \hat{S}(j,n) e^{2\pi j nk}
$$

and

$$
\hat{X}(k) = \sum_{j=0}^{N-1} \hat{TT}(j,k)
$$

3. Illustrations

The contour plots of the amplitude spectrum of the modified TT-transform have been shown in Figs. 2–5, for different values of the scaling parameter ‘c’. The analyzed signal (Fig. 1) is a mixture of sinusoid of frequency order 2, having amplitude 1. The transients taken are of the order of 10 and 20, respectively, each having amplitude 1. The first transient occurs in between the sampling points 100 and 200 and the second lies in between the points 300 and 350, respectively.

Fig. 1. Original time domain signal consisting of sinusoidal content of frequency order 2 and two transients of frequency order 10 and 20 centered at time points 150 and 350, respectively.
and 400. Thus the center or peak points for the transients are positioned at 150 and 350 sampling interval, respectively. The contour plots of Fig. 3 and Fig. 4 has been obtained for c=0.9 and c=0.7 and it is observed that, when c < 1, better frequency localization is obtained. For c=1.1, the window is squeezed around the center points, thus giving good time localization, while the contours are scattered over the frequency axis, which indicates poor spectral resolution level.

4. Applications to PQ disturbance

The proposed technique is used to analyze the power signal distortion in a realistic power network simulated by power system block set supported by MATLAB software. The sampling rate for the collection of power quality data is taken to be equal to 3.2 kHz. The time–time contours obtained by modified TT-transform and the magnitude of change are shown in Figs. 6–13 for different power quality patterns. The magnitude of change is derived from the TT transform resultant matrix by picking the respective variables are computed. Nodes at layer-3 represent the kernel function used in SVM. Layer-4 is the output layer. The signal propagation and the basic functions in each layer are described as follows:

Layer 1(input layer): No computation is done in this layer. Each node in this layer, which corresponds to one input variable, only transmits input values to the next layer directly. No computation takes place in this layer and the output of this layer o1 is expressed as o1 = x, where x, i = 1, 2, ..., M are the input features to the SVNN. Here we have considered only two input features x1 and x2.

Layer 2 (membership layer): Each node in this layer is a Gaussian membership function that corresponds one label of one of the input variables in layer-1. In other words, the membership value, which specifies the degree to which an input value belongs to a fuzzy set, is calculated by using the formula

\[ \mu_i(x) = \exp \left( -\frac{(x - c_i)^2}{\sigma_i^2} \right) \]

where, \( c_i \) and \( \sigma_i \) are, respectively, the center and the spread of the Gaussian membership function of the \( i \)th term of the input variable \( x \).

Then centers of membership function corresponding to each individual non-stationary power signal disturbances will be updated by the evolutionary computing approach explained in Section 8.

The spread is calculated by using the formula

\[ \sigma_i = \sqrt{\frac{\|x - c_i\|^2}{N}} \]

\( N \) is the no of data points.

5. Structure of the SVM classifier

A generalized four-layered support vector neural network (SVNN) is shown in Fig. 14, which comprises of the input, membership function, kernel function, output layers. Layer-1 accepts input variables whose nodes represent input variables. In layer-2 the membership values whose nodes represent the terms of the respective variables are computed. Nodes at layer-3 represent the kernel function used in SVM. Layer-4 is the output layer. The signal propagation and the basic functions in each layer are described as follows:

Layer 1(input layer): No computation is done in this layer. Each node in this layer, which corresponds to one input variable, only transmits input values to the next layer directly. No computation takes place in this layer and the output of this layer o1 is expressed as o1 = x, where x, i = 1, 2, ..., M are the input features to the SVNN. Here we have considered only two input features x1 and x2.

Layer 2 (membership layer): Each node in this layer is a Gaussian membership function that corresponds one label of one of the input variables in layer-1. In other words, the membership value, which specifies the degree to which an input value belongs to a fuzzy set, is calculated in layer-2.

The Gaussian membership function is defined as:

\[ \mu_i(x) = \exp \left( -\frac{(x - c_i)^2}{\sigma_i^2} \right) \]

where, \( c_i \) and \( \sigma_i \) are, respectively, the center and the spread of the Gaussian membership function of the \( i \)th term of the input variable \( x \).

Then centers of membership function corresponding to each individual non-stationary power signal disturbances will be updated by the evolutionary computing approach explained in Section 8.

The spread is calculated by using the formula

\[ \sigma_i = \sqrt{\frac{\|x - c_i\|^2}{N}} \]

\( N \) is the no of data points.
The center $c_j$ and spread $\sigma_j$ are fed as inputs to the kernel layer.

Layer 3 (kernel layer): In this layer the low dimensional input space is transformed into a high dimensional feature space. The reason behind this transformation lies in Cover's theorem. Cover's theorem states that "A complex pattern classification
problem cast in a high dimensional space nonlinearly is more likely to be linearly separable than in a low dimensional space. In our SVM the following types of kernel functions, which are listed below.

a. **Radial basis function**

\[ k(p,q) = \exp \left( -\frac{||p-q||^2}{2\sigma^2} \right) \]  

b. **Mexican hat wavelet kernel**

\[ k(p,q) = \frac{1}{\sqrt{2\pi\sigma^2}} \left( 1 - \frac{||p-q||^2}{\sigma^2} \right) e^{\frac{||p-q||^2}{\sigma^2}} \]  

The average value of the calculated spread in layer 2 is passed as one of the parameters \( p_\alpha \) of the kernel function. The other parameter of the kernel function is the center ‘\( q_\alpha \)’, which is a constant and can be assigned any value for achieving better classification.

Fig. 8. (a) Signal with voltage swell (b) magnitude of change (c) its time–time contour plot.

Fig. 9. (a) Signal with 10 Hz voltage flicker (b) magnitude of change (c) its Time-Time contour plot.
Layer 4 (output layer): In this layer the support vectors are obtained by solving the constrained quadratic programming problem. In this layer the support vector machine algorithm for classification has been applied.

6. SVM framework

Support vector machine is a very useful technique for data classification and regression problems. In most cases, the searching
of the suitable hyperplane in an input space is too restrictive to be of practical use. A simple intuition about the SVM can be obtained in the case of a linear classifier. Let a set $S$ of labelled training patterns $(u_1,v_1), \ldots, (u_l,v_l)$, each training point $u_1$ belongs to either of two classes with label $y\in\{1,-1\}$, for $i=0,1,\ldots,l$. In this case the SVM separates the data with a hyper plane described by $W^T U + b = 0$ that leaves the maximum margin between the two classes. Here, $W$ is a set of weight vector and $U$ is a vector consisting of input data. The task of maximizing the margin is equivalent to minimizing an upper bound on the generalization error of the classifier.

The hyper plane $W^T U + b = 0$ separates the data if $f(W^T U + b) \geq 1$.

To find the optimal separating hyper plane, one needs to find the plane, which maximizes the distance between the hyper plane

![Fig. 12.](image12.png)

(a) Voltage signal with momentary interruption (b) magnitude of change (c) its time–time contour plot.

![Fig. 13.](image13.png)

(a) Voltage signal with notch (b) magnitude of change (c) its time–time contour plot.
and the closest positive and negative samples. Therefore the classification problem is equivalent to minimizing
\[
\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{l} \xi_i \right)
\]
subject to:
\[
f\left( w^T u + b \right) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (i = 1..l)
\]
where, \(C\) is the penalty constraint, \(\xi_i\) \((i=1,...,l)\) are the slack variables and \(h\) is a constant.

Choosing the exponent \(h=0\) in the above equation results in nonappearance of the slack variable \(\xi\). Here, \(h\) is chosen as 1.

This constrained optimization problem can be translated into a dual quadratic programming problem by Lagrangian multipliers.

The saddle point of the primal Lagrangian gives the solution to a quadratic programming problem subject to inequality constraints. The Lagrangian function is:
\[
L_p(W,b,\xi,\alpha) = \frac{1}{2} W^T W + C \left( \sum_{i=1}^{l} \xi_i \right) - \sum_{i=1}^{l} \alpha_i [y_i (w^T u_i + b) - 1] - \sum_{i=1}^{l} \beta_i \xi_i
\]
where \(x_i\) and \(\beta_i\) are the Lagrange multipliers.

By solving Eq. (20) for optimal saddle point \((w_0, b_0, \xi_0, \alpha_0, \beta_0)\) we can get the dual lagrangian
\[
L_d(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j u_i^T u_j
\]

The roots of the above equation give the optimized Lagrange multipliers \(\alpha\). The values for which \(\alpha\) is minimum and non-zero are the support vector points.

The linear classifier presented in Eq. (21) is inadequate for the classification of the data points because of the overlapping nature of the data points. So for non-linear classifiers the input vectors \(U \in \mathbb{R}^d\) are mapped into vectors \(Z \in \mathbb{R}^d\).

\[
U \in \mathbb{R}^d \rightarrow Z(u) = [\phi_1(u), \phi_2(u), ..., \phi_n(u)]^T \in \mathbb{R}^d.
\]

This mapping leads to a linear classifier in a high dimensional feature space creating a non-linear separating hyper plane.

For this, the scalar product \(U^T U\) in Eq. (17) is replaced by the scalar product \(Z^T Z = [\phi_1(u), \phi_2(u), ..., \phi_n(u)] [\phi_1(u), \phi_2(u), ..., \phi_n(u)]^T\) in a feature space.

This is further expressed by the kernel function,
\[
K(u_i, u_j) = z_i z_j = \phi(u_i) \phi(u_j)
\]

Any symmetric function \(K(u,v)\) in input space can represent a scalar product in feature space if
\[
\int K(u,v) g(u) g(v) du dv > 0, \quad \forall g \in L_2(\mathbb{R}^d),
\]

where \(g(.)\) is any function with a finite \(L_2\) norm in input space or a function for which \(\int |g|^2 dx < \infty\).

Corresponding features in a \(Z\)-space are the eigen-vectors of an integral operator associated with \(K\),
\[
\int K(u,v) \phi_i(u) du = \lambda_i \phi_i(u)
\]
where, \(\lambda_i\) are the eigen values.

The kernels functions considered in this paper are Gaussian RBF kernel given by
\[
k(p,q) = e^{-\frac{|p-q|^2}{\sigma^2}}
\]
and Mexican hat wavelet kernel
\[
k(p,q) = \frac{1}{\sqrt{2\pi}\sigma} \left( 1 - \frac{|p-q|^2}{\sigma^2} \right) e^{-\frac{|p-q|^2}{\sigma^2}}
\]

### 7. Simulation results with Mexican hat wavelet kernel

Energy, standard deviation, autocorrelation, mean, variance, and normalized values have been extracted as features from the TT-transform analysis of the non-stationary power signals. Among these features the standard deviation and normalized values were found to be the most distinguished features. This conclusion was made from the visual verification from the feature plot, with different combination of features. Feature vectors (hundreds of each disturbance) are given as the input to the SVM network. Figs. 15–17 depict classification of voltage sag, voltage flicker, voltage swell, impulsive transient, harmonics, voltage notch, sag with harmonics, momentary interruption, oscillatory transient and voltage flicker with harmonics using Mexican hat wavelet c function.

When the classification was performed by the SVM algorithm with either RBF kernel or Mexican hat wavelet kernel function, the following serious drawbacks were observed.

(1) In Figs.15 and 17, the support vector boundaries of different power signal data patterns are overlapping with each other.

---

**Fig. 14.** Four-layered support vector network (SVN).

**Fig. 15.** Classification of voltage swell, sag, flicker, impulsive transient using Mexican hat wavelet kernel.
This results in an increased number of support vectors. Hence chances of misclassification will be quite prominent. In addition to this neither the support vectors boundaries are smooth nor the power signal data pattern are compactly clustered.

(2) In Fig. 16 it is seen that the support vector boundary for harmonic data pattern could not be drawn due the fact that some feature vector of harmonic data pattern belong simultaneously to sag with harmonic and voltage notch. This occurs because the Euclidean distance between the center of the harmonic data pattern and its corresponding feature vectors could not be minimized to a great extent. To circumvent this problem, its center as well as the spread of the proposed kernel function needs to be further fine-tuned using modified immune optimization algorithm given in Eqs.(25) and (26), respectively, of the next section.

8. Modified immune optimization algorithm (MIOA) and simulation result

The natural immune system can protect from a large variety of antigens, and the first problem it must face is the generation of a repertoire of antibodies with sufficient diversity to handle various antigens. Compared with the earlier existing immune algorithms, in the modified immune algorithm, the concentration of antibodies is employed to scale the diversity of solutions, and the mutation rate of antibody’s clone is dynamically adjusted based on its concentration and fitness. Moreover, this algorithm defines the clusters of antibodies, in which the elitist ones are always retained as the memory set, while other antibodies are replaced by fresh individuals Fig. 18 describes the flow chart for the modified artificial immune algorithm for center updation.

In our work each data point is considered as the antigen and the randomly selected center is taken as the antibody. The center (antibody) of the individual non-stationary power signal disturbances is updated by the following equation

\[ c_j = c_i + \alpha N(0,1) \]  

where

- \( c_i \) Initial random center from Eq. (15).
- \( c_j \) Mutated version of \( c_i \).
- \( \alpha \) Mutation rate of the antibody.
- \( N(0,1) \) Gaussian random variable with zero mean and standard deviation.

The mutation rate of the antibody is expressed as

\[ \alpha = \exp \left( -\beta \times \left( \gamma \frac{\text{concentration}(c_i)}{\text{mean(affinity)(x, c_i))} + (1 - \gamma) \times f^* \right) \right) \]  

where

- \( \beta \) Decay of mutation rate.
- \( f^* \) Fitness of the individual normalized in [0, 1].
- \( \gamma \) and \( (1 - \gamma) \) Are the two weighted parameters for the concentration and fitness, respectively.
- Concentration A measure of diversity.

Concentration of \( (c_i) \) is defined as follows:

\[ \text{concentration}(c_i) = \frac{\text{max(affinity)}}{\text{mean(affinity)(x, c_i))}} \]  

where, \( \text{max(affinity)} \) is the maximum of the affinities among all the antibodies, and \( \text{mean(affinity)(x, c_i))} \) is the mean of all the affinities between \( c_i \) and other antibodies.

Affinity is the Euclidean distance between two antibodies. The Euclidian distance is calculated according the formula

\[ \text{Distance} = ||x_i - c_i|| \]  

It can be seen from the definition, the concentration is based on the affinities (similarity) among the antibodies, and it is a measure of their diversity. If an antibody is relatively close to others in the search...
space, its concentration is higher, and the mutation rate ‘α’ of the clone is also higher. Hence, we expect this parent antibody to generate mutated clones with sparse locations and fitness to cover more domain space so that the solution diversity is maintained. The mutation rate of each antibody’s clone is a weighted value of the inverse of the concentration and fitness. If the fitness of the antibody is relatively high, its mutation rate is small. In other words, the mutation of the modified algorithm can achieve an appropriate trade-off between the exploration and exploitation in the search space.

Figs. 19–21 show classification of voltage sag, voltage flicker, voltage swell, impulsive transient, harmonics, voltage notch, sag with harmonics, momentary interruption, oscillatory transient and voltage flicker with harmonics using Mexican hat wavelet kernel function with MIOA.

The parameter used in the simulation of SVM classification for center updation is given below in Table 1. N is the number of input feature vectors for each of the power signal disturbances.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size N</td>
<td>100</td>
</tr>
<tr>
<td>Number of clones Nc</td>
<td>10</td>
</tr>
<tr>
<td>Decay of mutation rate β</td>
<td>10</td>
</tr>
<tr>
<td>Weighted parameter</td>
<td>0.9</td>
</tr>
</tbody>
</table>

$N_c$ is total no of clusters.

The proposed SVM algorithm with Mexican hat wavelet kernel functions when blended with modified immune optimization algorithm (MIOA) has shown superior performance in terms of distinct and smooth SVM clusters, probability of misclassification is reduced to large extent and the problem of drawing the support vector boundary for harmonic power signal data pattern was successfully overcome when compared to Mexican hat wavelet kernel function.

Fig. 22 describes the center updation of normalized value using Mexican hat wavelet kernel with modified immune optimization.
algorithm comparing the random center (old center) and the updated center (new center). The later one showing better convergence and classification of non-stationary power signal waveforms.

This paper has proposed a new approach for detection, localization and classification of non-stationary power signal waveforms using modified TT-transform and SVM classifier. The modified TT-transform with a variable window as a function of frequency is used using modified TT-transform and SVM classifier. The modified TT-transform provides a complete characterization derived for classification. Unlike the wavelet transform based techniques, the TT-transform provides a complete characterization of accuracy.

9. Conclusion

This paper has proposed a new approach for detection, localization and classification of non-stationary power signal waveforms using modified TT-transform and SVM classifier. The modified TT-transform with a variable window as a function of frequency is used to generate time–time distributions and significant features were derived for classification. Unlike the wavelet transform based techniques, the TT-transform provides a complete characterization of power signal disturbances. The support vector learning mechanism based on wavelet kernel provides a structural framework to extract support vectors from the input patterns. The choice of kernel function had a great influence in reducing the number of support vectors. The number of support vectors and the compactness of different clusters are further enhanced using modified immune optimization algorithm. As it can be inferred from Table 4, the proposed classifier outperforms earlier proposed classifiers in terms of accuracy.

References


Birendra Biswal received the Master of Engg. Degree from University College of Engineering, Burla, India in 2001 and Ph.D. from Biju Patnaik University of Technology, India. Presently working as Professor and Chairman Post Graduate Studies in GMR Institute of Technology, India. His primary research interest focuses on Digital Signal Processing, Adaptive Signal Processing, Statistical Signal Processing, Image Processing, Data Mining, System Identification and Machine Intelligence.

M. K. Biswal is an Assistant Professor in Silicon Institute of Technology, Bhubaneswar, India. He obtained his B. Tech. in Electronics and Telecommunications engineering from Silicon Institute of Technology. His primary research interests are applications of Signal Processing, and Machine Intelligence techniques for solving engineering problems.

S. Mishra (M’97-SM’04) received the B.E. degree from the University College of Engineering, Burla, Orissa, India, in 1990, and the M.E. and Ph.D. degrees from Regional Engineering College, Rourkela, Orissa, in 1992 and 2000, respectively. In 1992, he joined the Department of Electrical Engineering, University College of Engineering, as a Lecturer and, subsequently, became a Reader in 2001. Presently, he is an Associate Professor with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India. His current research interests include soft computing applications to power system control, power quality, and renewable energy. Dr. Mishra has been honoured with many prestigious awards, such as INSA Young Scientist Medal in 2002, INAE Young Engineer’s Award in 2002, and recognition as the DST Young Scientist in 2001–2002. He is a fellow of the Indian National Academy of Engineering, New Delhi, the Institution of Engineering and Technology, London, U.K., and the Institution of Electronics and Communication Engineering, India.

P. K. Dash Ph.D., D.Sc., SMIEEE, FNAE. He is currently working as Director (Research and Consultancy) in Sikha “O” Anusandhana University, Bhubaneswar, India. He is a visiting professor to USA, Canada, Malaysia & Singapore. Under his Supervision more than 40 research scholars have completed their PhD successfully. His area of research includes Power System, Control Engineering, Data mining, Intelligent Signal Processing, Soft Computing, Adaptive Signal Processing & Advanced Signal Processing.