Classification of No stationary Power Signals using Support Vector Machine and Extreme Learning Machine

Itishree Panda
M .Tech Scholar, EEE Department
Centurion University of Technology,
Bhubaneswar

Satyasis Mishra, PhD
HOD,ECE Department
Centurion University of Technology,
Bhubaneswar

ABSTRACT
The classification of nonstationary signals in a noisy environment is a difficult task. In this paper a modified version of S-Transform technique has been proposed for classification of power signal disturbances. The S-Transform is a signal processing technique which is used for visual localization, detection, pattern classification. S-Transform has good ability in gathering high frequency signals and suppressing the lower frequency signal. The S-Transform has been used to extract features from the nonstationary power disturbance signals.The extracted features are fed as the input support vector machine classifier for power signal disturbance pattern classification. To enhance the pattern classification accuracy the extreme learning classifier has been proposed and comparison results has been presented.

Keywords
SVM, power signals, S-Transform, STFT,Wt

1. INTRODUCTION
The power signal disturbances happen due to the use of electronic switching devices, power interruptions, capacitor switching and circuit faults. These devices introduce variations in the phase, frequency and amplitude of the power system signal. The signal processing techniques such as Discrete Fourier Transform (DFT), Short Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), S-Transform etc. are applied to nonstationary power disturbances signals for detection, localization and feature extraction and provides a powerful framework for feature extraction.DFT is insensitive to nonstationary signals, it shows only frequency spectrum not time information. STFT plays an important role by using a fixed window with signal, but fails to give variable resolution. Wavelet Transform (WT) [6] uses a variable window which gives good frequency resolution and poor time resolution at low frequency and good time resolution and poor frequency resolution at high frequency. The Wavelet Transform is sensitive to noise and does not retain the absolute phase information. So STFT and WT suffers from a trade-off between time and frequency resolutions.

To achieve absolute phase information and improved resolution, S-Transform is used which combines the good features of STFT [2, 3] and WT. S-Transform [8] uses a window which is inversely proportional to the frequency and a Fourier Kernel. In case of S-Transform, it uses a Gaussian window and provides variable resolution. To have good frequency resolution we have modified the window width of S-TransformFurther the modified S-Transform is applied to power disturbance signals for feature extraction. The extracted features are fed as input to the support vector machine classifier for classification. From the result it is found that it takes more computational time and complex calculation for classification. In this paper we have proposed the extreme learning machine algorithm in which simple mathematical calculations are involved and the comparison results have been usedThis paper presents four sections such as section II presents the derivation of S-Transform, Section III presents the proposed algorithm and the reference and conclusion is followed by section IV and V respectively.

2. S-TRANSFORM
The advantage of S-Transform [8] is that it preserves the phase information of the signal, and also provides a variable resolution similar to wavelet transform [3,6]. The standard S-Transform of a signal *x(t)* is given by a convolution integral [8] as

$$S(t, f) = \int_{-\infty}^{\infty} x(\tau) \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\tau)^2}{2\sigma^2}} .e^{-j2\pi f\tau}\, d\tau \qquad (1)$$

where f is frequency , t and τ are time variables.The standard deviation "σ "(dilation parameter or width of window) is a function of frequency f, and in normal S-transform is define as

$$\sigma = \frac{1}{|f|} \qquad (2)$$

The window provides good localization in the frequency domain for low frequencies while providing good localization in time domain for higher frequencies. The disadvantage of the current algorithm is the fact that the window width is always defined as a reciprocal of the frequency. A significant improvement of S-Transform can be realized by defining the standard deviation of the window as

$$\sigma = \frac{1}{|f|^{\gamma}} \qquad (3)$$

resulting in a modified S-Transform as

$$S(t, f) = \frac{|f|^{\gamma}}{k\sqrt{2\pi}} \int_{-\infty}^{\infty} x(\tau)\ e^{\frac{-(t-\tau)^2 f^{2\gamma}}{2k^2}} .e^{-j2\pi f\tau}\, d\tau \qquad (4)$$

Where, γ and α control the width of the window, the equation (4) can be written as

$$S(t,f)=\frac{|f|^{\gamma}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x(\tau)\ e^{\frac{-(t-\tau)^2 f^{2\gamma}}{2}}.e^{-j2\pi f\tau}\,d\tau \quad (5)$$

The S-Transform can be written as a convolution of two functions over the variable "t"

$$S(\tau,f)=\int_{-\infty}^{\infty}p(t,f)g(\tau-t,f)dt \quad (6)$$

Or $S(\tau,f)=\ p(t,f)*g(\tau,f)$ (7)

Where $p(\tau,f)=x(\tau)e^{-j2\pi f\tau}$ (8)

And $g(\tau,f)=\frac{|f|}{\sqrt{2\pi}}e^{\frac{-\tau^2 f^2}{2}}$ (9)

Let $B(\alpha, f)$ be the Fourier transform (from τ to α) of the S-Transform $S(\tau, f)$. By the convolution theorem the convolution in the τ (time) domain becomes a multiplication in the α (frequency) domain:

Now defining

$$B(\alpha,f)=P(\alpha,f)G(\alpha,f) \quad (10)$$

Where $P(\alpha, f)$ and $G(\alpha, f)$ are the Fourier transform of $p(\tau, f)$ and $g(\tau, f)$, so we can write

$$B(\alpha,f)=X(\alpha+f)\ e^{-\frac{2\pi^2\alpha^2}{|f|^2}} \quad (11)$$

Thus S-Transform is the inverse Fourier Transform of the above equation

$$S(t,f)=\int_{-\infty}^{\infty}X(\alpha+f)\ e^{\frac{-2\pi^2\alpha^2}{|f|^2}}.e^{+j2\pi\alpha\tau}\,d\alpha \quad (12)$$

and the modified S-Transform is given by

$$S(t,f)=\int_{-\infty}^{\infty}X(\alpha+f)\ e^{\frac{-2\pi^2\alpha^2}{|f|^{2\gamma}}}.e^{\,j2\pi\alpha\tau}\,d\alpha \quad (13)$$

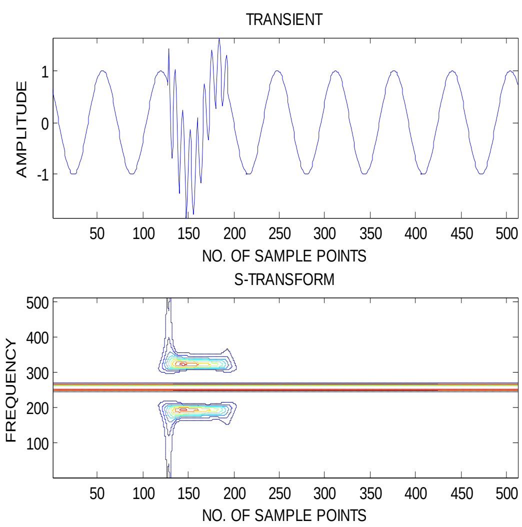

Fig .1 Localization of Transient signal using S-Transform

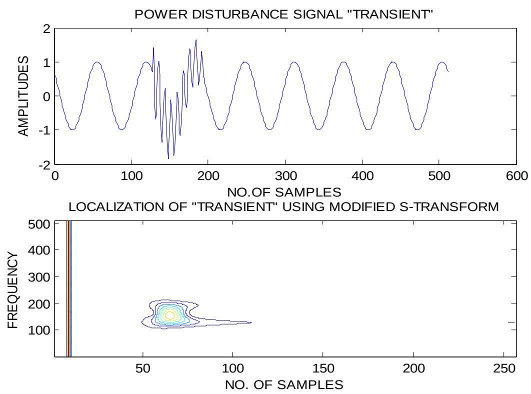

Fig .2 Localization of Transient signal using Modified S-Transform

3. PROPOSED METHOD

3.1 Support Vector Machine

For non-separable data, we will have alternative of mapping of data into higher dimensional space, as

$$x_i * x_j = \Phi(x_i).\Phi(x_j) \quad (14)$$

This higher dimensional space is called a feature space and then the kernel function is given by

$$\Phi(x_i).\Phi(x_j)\rightarrow K(x_i,x_j) \quad (15)$$

So the kernel is therefore the inner product between mapped pairs in the feature space.
Therefore the RBF kernel is given by

$$K(x_i,x_j)=e^{-\|x_i-x_j\|^2/2\sigma^2} \quad (16)$$

And this kernel must satisfy Mercers condition, for any $g(x), \int g^2(x)dx < \infty$ is finite and

$$\int K(x_i, x_j)g(x_i)g(x_j)dx_i dx_j \geq 0 \quad (17)$$

If a kernel which does not satisfy Mercer's condition, there may exist data such that the Hessian matrix is indefinite, and for which the quadratic programming problem will have no solution. The point of estimates of the weights $W = \{w_j\}_{j=0}^{N}$ needed by SVM[12], such that the linear combination of the base functions $\Phi = \{K(x_i, x_j)\}_{i,j=1}^{N}$ fits the training targets closely. At the same time, it reduces the complexity of computation by forcing the majority of the weights to zero.The SVM predictions is given by

$$y_i(x, w) = f_{svm}(x, w) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \quad (18)$$

$$y(x, w) = \Phi(x)w \quad (19)$$

And $w = [w_0, w_1, w_2, ...w_N]^T$

Where $K(x, x_i)$ is a RBF kernel function and it is the key factor in SVM to satisfy the "Mercer Condition". And Φ is an augmented kernel matrix and is given by

$$\Phi = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_1, x_1) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_N, x_1) & K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}$$

This matrix is formed by all the basis functions evaluated at all the training points that is with the RBF kernel functions.

For pattern classification problem, SVM we have to maximize

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i y_i \alpha_j y_j K(x_i, x_j) \quad (20)$$

Subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ where α_i, α_j are the hyper parameters and $i = 1, 2, ...N$

The decision function is given by

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + w_0 \right) \quad (21)$$

Some advantages of SVM are (i) it smoothly handles the non linear problems (ii) good prediction accuracy and involved simple mathematical calculations. SVM includes VC dimension and structural risk minimization for classification accuracy.

3.2 Extreme Learning Machine

The conventional SVM cannot be used in regression and multiclass classification applications directly. This paper shows that the extreme learning machine (ELM) algorithm works for single-hidden-layer feed forward networks where the hidden layer weights need not be updated.ELM was originally developed for the single-hidden-layer feed forward neural networks

$$f(x) = h(x)\beta \quad (22)$$

Where h(x) is the hidden-layer output corresponding to the input sample **x** and **β** is the output weight vector between the hidden layer and the output layer. One of the salient features of ELM is that the hidden layer need not be tunedare the initial starting values.

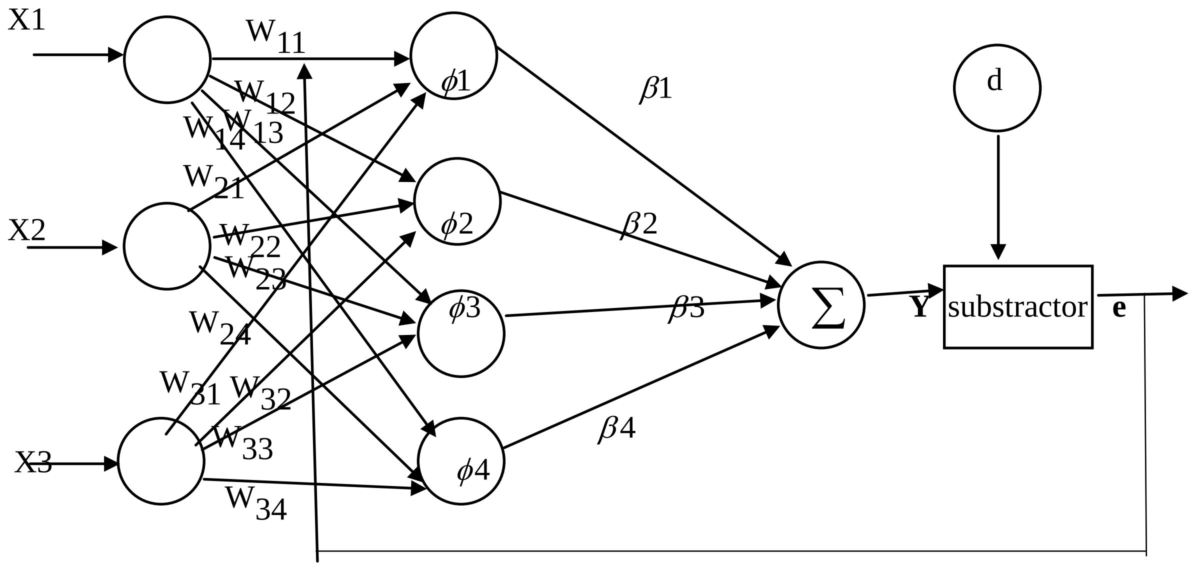

Fig.3 Single Layer Feed forward Neural Network

Given a set of N training dataset $D = (x_i, t_i), i = 1$ to N with each x_i is a d-dimensional vector and t_i is the expectation output. The output function of ELM[15] with L hidden neurons is represented by

$$f(x) = \sum_{k=0}^{L} \beta_k h_k(\theta_k; x) = h(\Theta; x)w \quad (23)$$

where $h(\Theta; x) = [1, h_1(\theta_1; x), \ldots\ldots., h_L(\theta_L; x)]$ is the hidden feature mapping with respect to input $x.\Theta = [\theta_1 \ldots\ldots \theta_L]$ are randomly generated parameters of hidden layer and w is the weight vector of all hidden neurons to an output neuron to be analytically analyzed. $h_k(\cdot)$ is the activation function of hidden layer. Equation (1) can be written as

$$\mathrm{H}\beta = \mathrm{T} \quad (24)$$

where H is the $\mathrm{N} \times (L+1)$ hidden layer feature-mapping matrix, whose elements are as follows:

$$\mathrm{H} = \begin{bmatrix} 1 & h_1(\theta_1; x_1) & \cdots & h_L(\theta_L; x_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & h_1(\theta_1; x_N) & \cdots & h_L(\theta_L; x_N) \end{bmatrix} \quad (25)$$

the i th row of H is the hidden layer's output vector for an instance x. Equation (2) is a linear system, which is solved by

$$\beta = \mathrm{H}^{\dagger}\mathrm{T}, \qquad \mathrm{H}^{\dagger} = (\mathrm{H}^T\mathrm{H})^{-1}\mathrm{H}^T \quad (26)$$

where $\mathrm{H}^{\dagger}$ is the Moore–Penrose generalized inverse [2] of matrix H .

In the proposed work features such as energy, standard deviation, autocorrelation, mean, variance, maximum and normalized values have been extracted from the nonstationary power signals.

The following disturbances have been considered for power signal clustering.

1. Transient
2. Harmonic
3. Notch
4. Sag
5. Spike
6. Sag+ Harmonic
7. Swell
8. .Flicker, Swell+ Harmonic

4. RESULTS

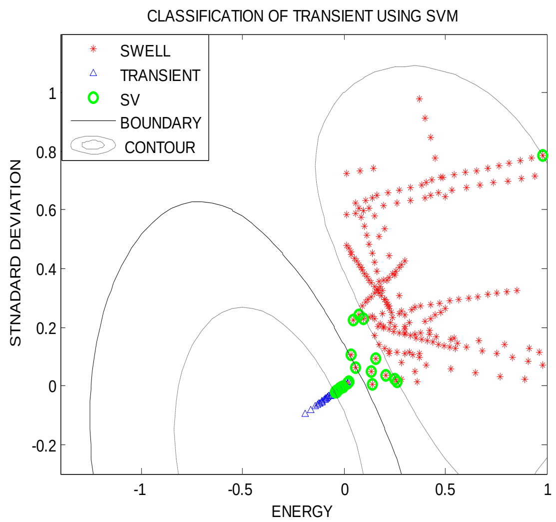

Fig .4 Classification of power signals using SVM algorithm

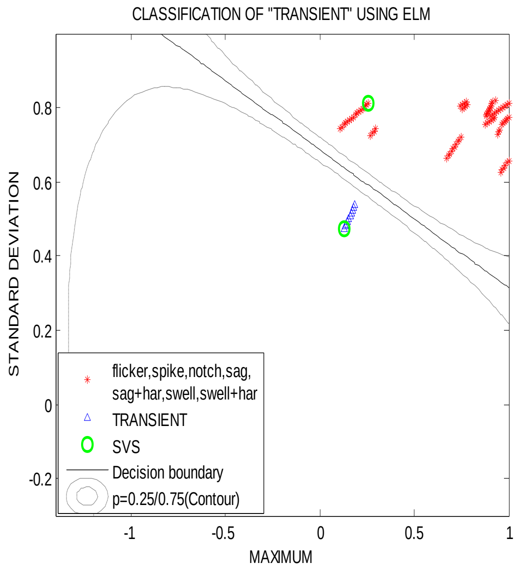

Fig .5 Classification of power signals using ELM

Fig.4 shows the classification of power signals using SVM algorithm and it is found that the power disturbance signals are classified but the cluster pattern visualization is not clear and the computational time is 21.0875 seconds. Fig .5 shows the classification of power signals using extreme learning machine classifier and the classification of the power signal is clear and the computational time taken is 10.3456 seconds.

Sl. No.	Power signal disturbances	Accuracy in percentage (%)	
		SVM	ELM
1	Transient	95.66	98.25
2	Harmonic	92.53	96.38
3	Notch	97.27	98.57
4	Sag	96.23	97.47
5	Spike	97.32	99.52
6	Sag+ Harmonic	94.21	98.45
7	Swell	94.84	97.12
8	Flicker, Swell+ Harmonic	96.88	98.43
%	Accuracy	95.62	98.02

The classification accuracy is shown in the table-1.

Table-1

5. CONCLUSION

The modified S-Transform has been applied to extract features and visual localization of nonstationary power disturbance signals. The modified S-Transform has good ability to localize the power signal waveforms better than the normal S-Transform and wavelet transform. The extracted features are fed as input to a support vector machine for pattern classification of power signal. To improve the classification extreme learning machine has been applied for pattern classification accuracy.

Extreme learning machine has shown higher pattern recognition accuracy in classifying power signal disturbances than the support vector machine classifier in terms of computational time. From the simulation results, it is found that the extreme learning machine classifier shows the better classification performance than the support vector machine classifier in power signal disturbance patterns classification.

6. REFERENCES

[1] L. Cohen, "Time-Frequency Distributions – A Review", *Proceeding of IEEE*, Vol.77, No.7, July 1989, pp.941-981.

[2] P. Rakovi, E. Sejdic, L.J. Stankovi and J. Jiang, "Time–Frequency Signal Processing Approaches with Applications to Heart Sound Analysis", *Computers in Cardiology*, Vol.33, pp.197–200, 2006.

[3] E. O. Brigham, "The Fast Fourier Transform And Its Applications", Prentice-Hall, Englewood Cliffs, New Jersey, 1988.F. S. Chen, "Wavelet Transform In Signal Processing Theory And Applications", National Defense Publication of China, 1998.

[4] I. Daubachies, "Ten Lectures On Wavelets", Philadelphia, PA: SIAM, 1992.

[5] S. Mallat, " A Wavelet Tour Of Signal Processing", London,U.K.:Academic,1998.

[6] Ingrid Daubechies, "The Wavelet Transform, Time–Frequency Localization and Signal Analysis", *IEEE Trans. On Information Theory*, Vol.36, No.5, pp.961–1005, 1990.

[7] R. Michael Portnoff, " Time–Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis*", IEEE Transactions On Acoustics, Speech, And Signal Processing*, Vol.Asp–28, No.1, pp.55–69, 1980.

[8] P. K., Dash, B. K., Panigrahi, & G. Panda, (2003). Power quality analysis using S-transform. IEEE Transactions on power delivery, 18(2), 406–411.

[9] Dash, P. K., Panigrahi, B. K., & Panda, G. (2003). Power quality analysis using S-transform. IEEE Transactions on power delivery, 18(2), 406–411.

[10] B. Biswal, P. K. Dash, S. Mishra, B. Biswal, P.K. Dash, S. Mishra, "A Hybrid Ant Colony Optimization Technique For Power Signal Pattern Classification", *Elsevier Science, Expert Systems With Applications*,Vol.38, No.5, pp.6368-6375, 2011.

[11] R. G., Stock well, L. Mansinha, & R. P Lowe,., (1996). Localization of the complex spectrum: The S-transform. IEEE Transactions on Signal Processing, 44(4), 998–1001.

[12] V. Vapnik, "The Nature of Statistical Learning Theory". *Springer Verlag*, New York, 1995.

[13] V. Vapnik, S. Golowich, and A. Smola, "Support Vector Method For Function Approximation, Regression Estimation, and signal processing". *Advances in Neural Information Processing Systems*, Vol.9, pp.281–287, 1996.

[14] M.E. Tipping. "Sparse Bayesian Learning And The Relevance Vector machine". *J Mach Learn Res*. Vol.1, pp.211–244, 2001.

[15] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine:Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, Dec. 2006.

[16] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learningmachine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern., Part B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[17] E. Soria-Olivas, J. Gomez-Sanchis, J. D. Martin, J. Vila-Frances, M. Martinez, J. R. Magdalena, and A. J. Serrano, "BELM: Bayesian extreme learning machine," *IEEE Trans. Neural Netw.*, vol. 22, no. 3,pp. 505–509, Mar. 2011.