Application of Pattern Recognition Method in Classifying Power System Transient Disturbance

Noraliza Hamzah, Fahteem Hamamy Anuwar, Zuhaina Zakaria
Faculty of Electrical Engineering,
Universiti Teknologi MARA, Shah Alam, Selangor, MALAYSIA
zuhaina@ieee.org

Abstract: - Power system transient can cause serious damage to main power system apparatus and sensitive loads. There are many causes of power system transient including capacitor bank switching, switching of large inductive loads and lightning. This paper discusses the application of pattern recognition method, namely Support Vector Machine (SVM) to classify the cause of transient disturbance in power system. Two types of feature extractions are applied to provide the inputs to the SVM, i.e. the minimum and maximum peak voltage values and the wavelet energy level of the transients. The IEEE 30 bus system is modeled using the Power System Computer Aided Design (PSCAD) software to generate different type of transient data caused by capacitor switching and lightning. Feature extraction is performed using discrete wavelet transform (DWT) analysis. The results showed that the performance of the feature extraction using maximum and minimum peak voltage values is superior (80%) as compared to the wavelet energy (54%) to classify the cause of the transient.

Key-Words: - Power quality, transient, discrete wavelet transform, support vector machine, radial basis function

1 Introduction
Transient are sub-cycle deviations from the normal voltage or current waveform. A number of transient causes are lightning strokes, switching actions in the distribution or transmission system, self-clearing faults or faults clear by current limiting fuses, and the switching of end-user equipment. Transient occurs in a very short duration (microseconds to milliseconds) and can be of large magnitude [1]. Low frequency transient due to capacitor bank energizing have received the most attention in the power quality literature due to their adverse impact on the adjustable speed drives [1-3].

Therefore, in order to improve the quality of power system, it is desirable to develop method for identifying the causes of power quality disturbance which include transient. Li et. al [4] has employed SVM for classification of various cases of line energization in the category of safe or unsafe based upon the peak value of overvoltage at the receiving end of line. Przemyslaw et. al [5] used a multi class SVM to classify the disturbance whilst Tong et. al [6] employed an integrated SVM and the one-versus-one (OVO) based SVM which can process the multiple classification in order to reduce the processing time for PQ detection. Based on the findings in [4-9], this paper proposes the application of SVM to classify the cause of transient either by capacitor switching in power distribution system or due to lightning.

This paper proposes two types of feature extractions to be employed as the inputs of the SVM. The accuracy between these two feature extractions are analyzed and compared.

2 Theories
This section describes theories of wavelet and support vector machine that are used in this paper to classify the cause of transients.

2.1 Wavelet
Wavelet transform (WT) is a mathematical technique used for many application of signal processing. Wavelet is much more powerful than conventional methods in processing the stochastic signal because of analyzing the waveform time-scale region [10]. This paper used the applicability of the Discrete Wavelet Transform (DWT) to the analysis of power system transient [10-11]. The DWT is defined as:

$$DWTx[m,k] = \frac{1}{\sqrt{a_0^n}} \sum_n x[n] \frac{k - nb_s a_o^n}{a_o^n}$$  

(1)
where \( g(.) \) is the mother wavelet whilst \( a=a_o^m \) and \( b=b_o a_o^m \) are the scaling and translation parameters of integer ‘m’ respectively. By careful selection of \( a_o \) and \( b_o \), a family of dilated mother wavelets can be generated which has an orthonormal basis with respect to the \( L^2(\mathbb{R}) \) space. By simple interchange of variables \( n, k \) and rearrangement of the DWT in (1), we get:

\[
DWTx[m,n] = \frac{1}{\sqrt{a_o^m}} \sum_k x[k] g[a_o^m n - b_o k] \quad (2)
\]

Upon closer observation of this equation, we notice that it is similar to the equation for digital FIR filters;

\[
y[n] = \frac{1}{c} \sum_k x[k] h[n-k] \quad (3)
\]

By comparing equations 2 and 3, the impulse response of the filter in the DWT equation is

\[
g[a_o^m n - b_o k]. \quad (4)
\]

If \( x[n] \) is the discrete time signal to be analyzed then the signal is decomposed into two signals which contain the high and low frequency components as in (5) and (6) respectively.

\[
d_i[n] = \sum_k x[k] h[2n-k] \quad (5)
\]

\[
c_i[n] = \sum_k x[k] g[2n-k] \quad (6)
\]

The high frequency component \( d_i[n] \) is the detailed version of the signal. The low frequency components, \( c_i[n] \), is further split to get the other details of input signal. By using this technique, any mother wavelet can be implemented although certain wavelets generate symmetric filter coefficients which have certain desirable properties.

In multi resolution analysis (MRA), wavelet and scaling functions are used as building blocks to decompose and construct the signal at different resolution levels [12]. The main goal of MRA is to developed representations of a signal at various levels of resolution. MRA is composed of two filters in each level which are low pass filter (LPF) and high pass filter (HPF).

A set of orthogonal scaling function \( \phi(t) \) and wavelet function \( \psi(t) \) can divide function space into a series of orthogonal high frequency and low frequency spaces. The translated and scaled version of the wavelet used in MRA is time frequency picture of the decomposed signal \( f(t) \). The original signal \( f(t) \) can be decomposed to:

\[
f(t) = \sum_k c_j(k) \phi_k(t) + \sum_{j=1}^k \sum_{k} d_j(k) \psi_k(t) \quad (7)
\]

Where \( j \) is the level number of the wavelet decompositions. \( c_j, d_j \) are the weighted sequences of the space projection of \( f(t) \), respectively.

### 2.2 Support Vector Machine

Support vector machine is a very useful technique for data classification and regression problems. SVM have been used in many pattern recognition and regression estimation problems and have been applied to the problems of dependency estimation, forecasting and constructing intelligent machines [12].

Let \( n \)-dimensional inputs \( x_i (i=1,2,\ldots,M) \), \( M \) is the number of samples belong to class -1 or class -2 and associated to labels \( y_i=1 \) for class -1 and \( y_i=-1 \) for class -2. For linear separation data a hyper plane \( f(x)=0 \) which separates the data can be determined as shown in (8).

\[
f(x) = w^T x + b = \sum_{j=1}^n w_j x_j + b = 0 \quad (8)
\]

where \( w \) is an n-dimensional vector and \( b \) is a scalar.

The vector \( w \) and the scalar \( b \) determine the position of the separating hyper plane. This separating hyper plane satisfies the constraints \( f(x_i) \geq 1 \) if \( y_i=1 \) and \( f(x_i) \geq -1 \) if \( y_i=-1 \) and this results in

\[
y_if(x_i) = y_i(w^T x_i + b) \geq +1 \quad (9)
\]

for \( i=1,2,\ldots,M \).

Considering noise with the slack variables \( \xi_i \) and the error penalty \( C \), the optimal hyper plane can be found by solving the following convex quadratic optimization problem as given by,

\[
\frac{1}{2} ||w||^2 + C \sum_{j=1}^M \xi_i \quad (10)
\]

Equation (10) is subject to,

\[
y_i(w^T x_i + b) \geq 1 - \xi_i \quad (11)
\]
for \( i = 1,2, \ldots \ldots M \) and \( \xi_i \geq 0 \) for all \( i \)

where \( \xi_i \) is the distance between the margin and the examples \( x_i \) lying on the wrong side of the margin. Applying the Kuhn-Tucker conditions for maximization into the equivalent Lagrange dual problem as given by,

\[
W(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,k=0}^{M} \alpha_i \alpha_k y_i y_k x_i^T x_k
\]  

(12)

which is subject to,

\[
\sum_{i=0}^{M} y_i \alpha_i = 0
\]  

(13)

for \( C \geq \alpha_i \geq 0, \ i = 1,2, \ldots \ldots M \)

The number of variables in the dual problems gives the number of training data. Denoting the optimal solution of the dual problem with \( \alpha^* \) and \( w^* \) the equality conditions in () holds for the training input-output pair \( (x_i, y_i) \) only if the associated \( \alpha^* \neq 0 \). In this case, the training example \( x_i \) is called support vector (SV). The number of SVs is considerably lower than the number of training samples making SVM computationally very efficient. The value of the optimal bias \( b^* \) is found from the geometry;

\[
b^* = -\frac{1}{2} \sum_{SV} y_i \alpha_i^* \left( s_1^T x_i + s_2^T x_i \right)
\]  

(14)

where \( s_1, s_2 \) are the arbitrary SVs for class -1 and class -2, respectively. Only the samples associated with SVs are summed because the other elements of the optimal Lagrange multiplier \( \alpha^* \) are equal to zero. The final decision function is given by

\[
f(x) = \sum_{SV} \alpha_i^* y_i \phi^T(x_i) \phi(x) + b^*
\]  

(15)

The unknown data sample \( x \) is then classified as,

\[
x \begin{cases} 
\text{class } -1, & \text{iff } f(x) \geq 0 \\
\text{class } -2, & \text{otherwise}
\end{cases}
\]

The nonlinear classification problems can also be solved by using SVM applying a kernel function. The classified data is mapped onto a high-dimensional feature space where the linear classification is possible. Using a nonlinear vector function,

\[
\phi(x) = \phi_1(x), \phi_2(x), \ldots, \phi_m(x)
\]

for \( m \gg n \)

To map the n-dimensional input vector \( x \) into the m-dimensional feature space, the linear decision function in dual form is given by,

\[
f(x) = \sum_{SV} \alpha_i^* y_i K(x_i, x)
\]  

(18)

Notice that in (), the inner products are used. A function that returns a dot product of the feature space mapping of the original data points is called a kernel function \( K(x,z) = \phi^T(x)\phi(z) \). The learning in the feature space does not require the inner products where a kernel function is applied. Using a kernel function, the decision function can be written as,

\[
f(x) = \sum_{SV} \alpha_i^* y_i K(x_i, x)
\]  

In this paper, Gaussian radial basis kernel function which gives the best results is selected. The radial basis kernel function is defined as,

\[
K(x,z) = \exp \left( -\frac{1}{2\sigma^2} \right)
\]  

(19)

where \( \sigma \) is the width of the Gaussian function.

3 Results and Discussions

The transient signals are obtained from the PSCAD simulation. The input signal is transformed into the time-frequency domain using DWT. When the SVM classifier is trained, the free regularization constant \( C \) and kernel function argument \( \sigma \) must be determined. Common method to tune free parameters is to use cross-validation technique. In this paper, after performing the cross validation test, it is shown that SVM gives better outputs when the value \( c=8 \) and \( \sigma=2 \). Radial Basis Function (RBF) has been selected as the kernel function.

As mentioned earlier, two types of transient is considered in this paper that is caused by capacitor switching and lightning. Figure 1a) and b) shows the waveform of transient which is caused by capacitor switching and lightning respectively.

Figure 2a) and b) shows the wavelet of transient caused capacitor switching and lightning respectively.
The fourth order Daubechies wavelet (Db4) was adapted to analyze the disturbance signal since it has the best similarity to power disturbance signals. The samples disturbance signals were decomposed with 10 levels MRA to ensure that all disturbance features are extracted in both high and low frequency spectrum.

Support Vector Machine is trained to determine and identify the support vectors which are the parameters that define the optimal hyper plane. Results at level D3 of db4 were selected as an input to SVM. The SVM input vectors that are the maximum and minimum peak value of the wavelet detail function of the wavelet are normalized within [0-1].

Normalization is done to eliminate the risk of numerical problems during the training process. The processed features are stored in \( m \) by \( n \) dimensional training matrix, in which \( m \) and \( n \) is input and output of the SVM respectively. A 2D SVM classifier is employed with +1 and -1 for the transient caused by the capacitor switching and lightning respectively. A total of 96 samples constitute the database with both samples of capacitor switching and lightning that are equally divided. The samples are defined into two classes with +1 for capacitor switching and -1 for lightning. In this experiment, about 70% samples are randomly selected to construct the training data set, whilst the remainder is used for testing. The proportion of each class sample in both dataset is equal. The results of the classification using SVM in which the minimum and maximum voltage peak values as the inputs are presented in Fig. 3. Fig. 3 shows that, the transient caused by lightning and capacitor switching is denoted by class +1 and -1 respectively. In this experiment, it has been found that the testing classification accuracy is 80%. These results indicate that the SVM is able to classify the cause of the transient effectively.

Results on the wavelet energy as the input of the SVM are presented in Fig. 4. Fig. 4 a) and b) shows the energies of 10 level decomposition for transients caused by capacitor switching and lightning respectively. As shown in Fig. 4, energy level 8 and 9 are more significant as compared to the other levels. These distinctive features are important for the performance of classifier. Hence, these energy levels are used as the inputs for SVM classification. The feature set are normalized between [-1,1] before presented to classifier to reduce space and enhance the performance of the classifier.
For classification using energy level 8 and 9 as the SVM inputs, a total of 72 samples constitute the database both of capacitor switching and lightning have been used. The samples are defined into two classes, with +1 and -1 for capacitor switching and lightning respectively. In this experiment, about 70% sample is randomly selected to construct the training dataset, whilst the remainder is used for testing. The proportion of each class sample in both dataset is almost equal. Fig. 5 shows that the classification for the cause of the transient and capacitor switching is denoted by class +1 and -1 respectively. In this experiment, it has been found that the testing classification accuracy is 54%. This results indicate that only 54% of the transient cause is correctly classified using the energy level as the SVM inputs.

4 Conclusions

In this paper, results attained using RBF kernel function of the SVM has proved that the method can be used to classify the causes of transient. Comparison made between the two types of feature extraction, the minimum and maximum of peak voltage values and the wavelet energy level reveal that, the former is superior in accuracy (80%) to classify the cause of transient. The RBF kernel has been identified as the best kernel to classify the causes of transient as either by capacitor switching or lightning in the power system distribution system.

Acknowledgments

The authors would like to acknowledge the Ministry of Science and Technology Invention for the grant to support the research and also the government of Malaysia for the scholarship awarded.

References:

[8]. Sheng-Wu Xiong, Ziao-XiaoNiu, Hong-Bing Liu, Support Vector Machines Based On Subtractive Clustering, Proc of the 4th Int. Conference on

[10]. Resende, J.W., Chaves, M.L.R., Penna, C., Identification if Power Quality Disturbances Using The MATLAB Wavelet Transform Toolbox, Universidade Federal de Uberlandia (MG)-Brazil
