International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:9, 2010

Fault Zone Detection on Advanced Series Compensated Transmission Line using Discrete Wavelet Transform and SVM

Renju Gangadharan, *Graduate Student Member*, *IEEE* , G. N. Pillai, and Indra Gupta

***Abstract*—**In this paper a novel method for finding the fault zone on a Thyristor Controlled Series Capacitor (TCSC) incorporated transmission line is presented. The method makes use of the Support Vector Machine (SVM), used in the classification mode to distinguish between the zones, before or after the TCSC. The use of Discrete Wavelet Transform is made to prepare the features which would be given as the input to the SVM. This method was tested on a 400 kV, 50 Hz, 300 Km transmission line and the results were highly accurate.

***Keywords*—**Flexible ac transmission system (FACTS), thyristor controlled series-capacitor (TCSC), discrete wavelet transforms (DWT), support vector machine (SVM).

I. Introduction

THE use of Flexible AC Transmission System (FACTS) devices has become quite popular in transmission systems. The Thyristor Controlled Series-Capacitor (TCSC) is one such device that has found usage in the transmission lines as it has enabled optimum use of the transmission line by allowing features like power control, dynamic compensation etc. The TCSC consists of a series capacitor in parallel with a series combination of a reactor and an anti-parallel connection of thyristors along with a protection feature in the form of a Metal Oxide Varistor (MOV). The MOV provides over-voltage protection to the series capacitor. However, the introduction of the TCSC has caused the protection schemes of transmission lines using traditional methods, to be difficult because they introduce sudden changes in the apparent impedance of the transmission line [1],inversion of voltage and current signals etc. [2,3]. Since, protection schemes, existing now, determine the state of the system through the voltage and current waveforms, the fault analysis has now essentially become a pattern recognition problem.

Different kinds of approaches have been proposed to find out the faulty zone. [4] uses the travelling wave approach for the protection of the series compensated line, [5] and [6] make use of the SVM and a combination of the SVM and DWT respectively and [7] make use of a scheme called the Decision Tree (DT) to find out the faulty zone.

Renju Gangadharan (renjumace@gmail.com), G.N. Pillai (gnathfee@iitr.ernet.in) and Indra Gupta (indrafee@iitr.ernet.in) are all with the Dept. of Electrical Engg. at the Indian Institute of Technology Roorkee, Roorkee-247667, India.

The above mentioned works, however, do not produce satisfactory results for section identification with a maximum overall accuracy of 93.917% has been reported in [6].

The method proposed here also uses both the SVM and the DWT, but a thorough analysis was done to find out the best features that can aid the process. The pre-requisite for the proposed method is that the type of fault viz. a-g, b-g, c-g, a-b-g, a-c-g, b-c-g, a-b-c-g, a-b, a-c and b-c must be found. A data window of 2 cycles corresponding to 200 samples was taken for the analysis. . The idea of using 2 cycles' information is to ensure that all the possible test cases can be incorporated. The data obtained after processing was given to the SVM for classification. The SVM algorithm classifies the data into two classes i.e. SVMs solve for two class problems. The section identification is for a particular type of fault, hence, a total of 10 SVMs are needed to determine the faulty section.

In this paper, the second section gives a brief insight into the wavelet transforms. The third section gives an account of the classification using SVM. The fourth section describes the system studied and the simulation process. The fifth section presents the feature extraction process. The sixth section describes the results and the last section concludes the work

II. Discrete Wavelet Transform

Before discussing about DWT, a brief introduction of continuous wavelet transform or simply the wavelet transforms should be made. The wavelet transforms are realised by two sets of functions called the scaling function and the wavelet function. The expressions for these are,

$$\varphi_{jk} = 2^{j/2}\varphi_{00}(2^j t - k) \quad (1)$$

$$\psi_{jk} = 2^{j/2}\psi_{00}(2^j t - k) \quad (2)$$

where φ represents the scaling function and ψ represents the wavelet function. φ_{00} represents the basis function for the scaling function and similarly, ψ_{00} represents the basis function for the wavelet function. It is also called the mother wavelet. The parameter *j* represents the resolution level and the parameter *k* is called the shifting parameter. The parameter *k* ranges from 0 to 2^{j-1} while the parameter *j* starts at 0 and ideally goes to infinity but, in practice it is limited by the sampling rate. Hence, the scaling and the wavelet functions are the scaled and shifted versions of their corresponding basis functions. However, the two scale relation in the wavelet theory helps in defining the wavelet and the scaling function at a particular resolution level purely in terms of the scaling

function of the next higher resolution level. The expressions for the scaling function and the wavelet function now turn out to be like this,

$$\varphi(t) = \sum_k h_0(k)\sqrt{2}\varphi(2t-k) \quad (3)$$

$$\psi(t) = \sum_k h_0(k)\sqrt{2}\psi(2t-k) \quad (4)$$

The coefficients h_0 and h_1 are considered as filter coefficients, h_0 corresponds to a low pass filter and h_1 corresponds to a high pass filter, thus, rendering the wavelet transforms to be considered as a filter operation.

Just like the discrete Fourier transform, the DWT can also be thought of as an operation that maps continuous time signals into a set of numbers. These coefficients are denoted by making use of the two scale relation as

$$c_j(k) = \sum_m h_0(m-2k)c_{j+1}(m) \quad (5)$$

$$d_j(k) = \sum_m h_1(m-2k)c_{j+1}(m) \quad (6)$$

The coefficient c is termed as the approximation coefficient; the term d is termed as the detail coefficient and m is the sample instant. We can clearly see from the expressions that the detail coefficients correspond to the high frequency coefficients and the approximation coefficients correspond to the low frequency coefficients. The inverse synthesis to obtain the original signal is given by,

$$f(t) = \sum_{k=0}^{2^{j-1}} c_{jk}\varphi_{jk}(t) + \sum_{j=J}^{\infty}\sum_{k=0}^{2^{j-1}} d_{jk}\psi_{jk}(t) \quad (7)$$

The DWT brings with it an important concept of Multi Resolution Analysis (MRA). It develops representation of signal in terms of wavelet and scaling functions in different frequency resolution levels. This can be ascertained from (5) and (6) i.e. the detail and approximation coefficients of a particular resolution level can be obtained from the approximation coefficients of the previous level. Hence, the MRA can be considered as a sequential combination of pairs of high pass and low pass filters to obtain the detail and approximation coefficients at each frequency resolution level. Fig. 1 shows the concept of the MRA where in the end we obtain the detail coefficients of all the resolution levels and the approximation coefficients of the last resolution level. This sort of analysis is ideal for fault analysis because it requires the higher frequency components to be examined and the detail coefficients correspond to the same only.

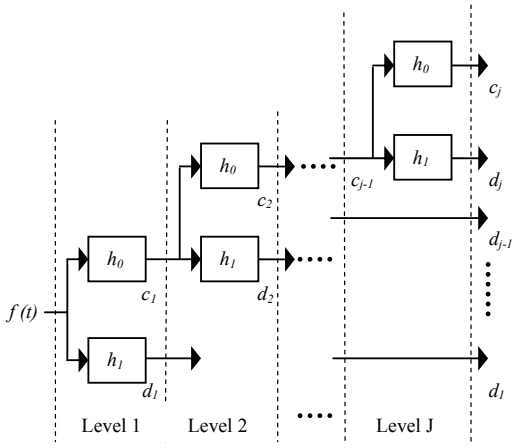

Fig. 1. Multi-Resolution Analysis using DWT

III. Classification using SVM

SVMs are a set of related supervised learning methods used for classification and regression. Since their advent they have become the most widely used tools for the classification of data. The SVM, like other classifiers, separates the data into two classes represented by labels. They classify a set of non-linear data by first exporting them to a higher dimensional feature space and then creating a hyperplane to separate them in the same feature space. Even though the given data might not be linearly separable, but the hyperplane generated in the higher dimension space is a linear one.

SVMs have been widely used in the image processing techniques because it can handle very large dimensional data quite easily as compared to other learning algorithms like the neural networks. The same attribute of the SVMs can be applied in the classification of faults on a transmission line as well, taking into considerations the number of inputs/features that would have to be provided. One more advantage that the SVMs have over the other learning algorithms is that it works on the principle of Structural Risk Minimization (SRM), which is rooted in the statistical theory. This technique involves the minimization of the upper bound (it is a sum of the training error rate and a term that depends on the Vapnik-Chervonekis (VC) dimensions [8]) of the generalization error [9, 10]. The other learning algorithms use another technique called the Empirical Risk Minimization (ERM) i.e. they tend to reduce the error on the training dataset. Neural networks are a prime example of ERM technique. On comparison, the SRM technique was found to give better generalization abilities i.e. to classify unseen data correctly.

Therefore, the main objective of an SVM is to create a hyperplane to classify a set of data into two classes so that the margin between the sets is a maximum. Consider a vector of data points denoted by x_i (i=1... m) which belong to either class-I or class-II, denoted by y (the class can be given a label, usually y=1 or y=-1). Now the equation of the separating hyperplane in any dimension will be given by

$$f(x) = w^T x + b = 0 \quad (8)$$

where w correspond to a weight vector having the same dimensions as the input vector x_i. The parameter b is a scalar. If a data point from the vector x_i makes f(x)>0 for a particular w and b, then that point belongs to (say) class-I and the data point that corresponds to f(x)<0, belongs to class-II. The parameters w and b determine the position of the hyperplane. The above discussed scenario was for a linearly separable data. However, in practice we seldom encounter such problems.

The maximization of margin between the two classes involves the maximization of a parameter known as the geometric margin. It is the distance of the separating hyperplane from a point and can be used as a parameter to formulate the problem statement for the SVM. The geometric margin is showed in Fig. 2. The geometric margin is equal to $1/\|w\|$ and using this, the problem statement can be stated as

$$y_i(w^T x_i + b) \geq 1 - \zeta_i$$
$$\zeta_i \geq 0, \qquad \text{for } i=1....m \quad (9)$$

The parameters ζ_i and C are employed as means to decide upon the rogue points i.e. the points that are misclassified. Usually if a point is misclassified during training, the value of the parameter ζ_i becomes greater than one and in order to discourage the classifier to do this, a penalty is introduced on that rogue point in the training example as represented by the second term in the objective function. The parameter C represents the trade-off between the optimal margin and the misclassification. The solution for the above expressions can be found out using the Lagrange multiplier method, the duality concept and the Karush Kuhn Tucker conditions and the final expression is obtained as,

$$\max Q(a) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j x_i^T x_j$$

subject to, $0 < \alpha_i < C$,

$$\sum_{i=1}^{m} y_i \alpha_i = 0, \qquad i=1.....m, \quad (10)$$

The number of elements in the vector x represents the number of inputs. However, the training happens only for those elements in the input vector for which the corresponding α_is are not zeroes. Those points for which the corresponding α_is are not equal to zero are called the support vectors (SVs) and hence, the name Support Vector Machine. The advantage of this phenomenon is that since, the SVs are always less in number than the total amount of input vectors, the training is always fast. The optimum value of the parameter b is given by

$$b^* = (\max_{\forall i, yi=-1} w^T x_i + \min_{\forall i, yi=1} w^T x_i)/2 \quad (11)$$

This expression can be thought off intuitively as being half of the distance between the nearest SVs of the either class.

On the application of the KKT conditions on the lagrangian equation of (9), the optimal value of the parameter w is obtained as

$$w^* = \sum_i \alpha_i y_i x_i \quad (12)$$

On substituting the optimum values of the parameters b and w in (8), the final expression of the decision hyperplane comes as,

$$f(x) = \sum_{SV} \alpha_i y_i x^T{}_i x + b^* \quad (13)$$

The summation should be done for the whole range of i but, since the value of the α_i will be non zero only for the SVs, hence only the SVs are considered. Any test point will be classified by this decision hyperplane into either class-I or class-II. With this, we can seen that the SVMs reduce the training effort by concentrating only on the SVs. Fig. 3 shows a decision hyperplane in a two dimension plane formed with the help of the SVs.

We can see from (10) that the only dependence of the algorithm on the input x is through the inner product $x_i^T x_j$. Suppose, if we introduce a function φ such that it maps the input vector x into a higher dimensional feature space, so that the data which was linearly inseparable in the current feature space is linearly separable in the higher dimensional feature space. That is,

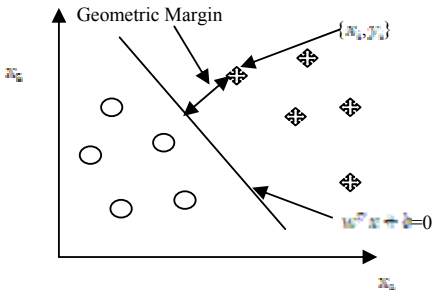

Fig. 2. Geometric margin of a point from the separating hyperplane

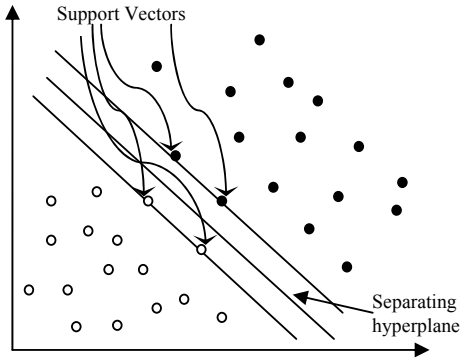

Fig. 3. Creation of separating hyperplane using support vectors

$$x \xrightarrow{map\ w.r.t.\ \phi} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \phi(x) \quad (14)$$

Therefore the inner product $x_i^T x_j$ in the previous lower dimensional feature space can be now represented as $\varphi(x_i)^T\varphi(x_j)$ in the higher dimensional feature space. However, this approach can make the computation very difficult as it would not be feasible to represent a higher dimensional feature vector like $\varphi(x)$, leave alone the inner product $\varphi(x_i)^T\varphi(x_j)$. This problem is overcome by the application of the kernel functions as they reduce the computational effort.

The inner product in the higher dimensional feature space i.e. $\varphi(x_i)^T\varphi(x_j)$ is replaced by a kernel function as

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (15)$$

To find out how a kernel function can reduce the computational effort, the following example can be of some help. Suppose that we have two input vectors *x & z*, such that $x,z \in R^{max}$ and we have a kernel function *K(x,z)* such that,

$$K(x,z) = (x^T z)^2 = (\sum_{i=1}^{n} x_i z_i)(\sum_{j=1}^{n} x_j z_j) \quad (16)$$

$$(\sum_{i=1}^{n} x_i z_i)(\sum_{j=1}^{n} x_j z_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i x_j)(z_i z_j)$$

$$\text{so, } (\sum_{i=1}^{n} x_i z_i)(\sum_{j=1}^{n} x_j z_j) = (\phi(x)^T)(\phi(z)) \quad (17)$$

From the above expressions it can be observed that the computational time to calculate the expression $\varphi(x)$ is in the order of n^2, where n is the dimension of x and z, whereas, the time required to compute the kernel function *K(x,z)* is in the order of n only. That's how a kernel function makes the computation of the inner product efficient

A query can be put forward as to whether there exist a function $K(x_i,x_j)$ which can act as kernel or can any function can act as a kernel. The validity of a kernel function is given by the *Mercer's Theorem* [11] which states that only positive definite functions can act as kernels. The kernel function in the SVM employed for this paper is radial basis function (RBF) given as,

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2) \quad (18)$$

where σ (sometimes mentioned as *g*) is the width of the RBF function.

IV. System and Simulation Studies

The system used for the simulation studies is a 400 kV, 50 Hz, 300 km transmission line having a TCSC module in the middle and two sources connected to its both ends, one acting as the actual power source and the other as an assumed infinite source. The relaying is done at the sending end i.e. before the TCSC module. All the components are realized using the PSCAD/EMTDC subroutines [12]. Fig. 4 shows the schematic diagram of the system considered.

In order to cover the maximum possible cases, certain parameters viz. the firing angle (FA) of the thyristors, the fault resistance (Rf), the fault inception angle (FIA) and the fault location were varied. To provide capacitive compensation to the transmission line, the TCSC is designed so as to provide a maximum compensation level at a FA of 150^0 and a minimum compensation level at an FA of 180^0. Usually, the TSCSs are designed to provide a compensation level of nearly 30-40% only.

Similarly, the fault resistance is varied within a range of 0 to 50 ohms; the FIA is varied within a range of 0^0 -180^0 , the fault location is selected at 5% -95% at an interval of 5% of the transmission line. Apart from the variation in the fault location, the number of variations in the FA, fault resistance and the FIA is 10, 4 and 5 respectively for each fault type. As the number of fault types is 10 (stated in section 1), the total amount of the variations or to say the total amount of variations come out to be 38000.

V. Feature Extraction

The three voltage waveforms were taken for extracting the features as they were the ones with appreciable characteristics for the purpose of section identification as compared to the current waveforms. The analysis was done on a-g fault first and the corresponding features resulting from its analysis were also taken for the other faults.

As mentioned earlier that DWT can be considered as a filter operation. Daubechies' wavelets with different orders have been utilized here for the analysis for e.g. *db5* implies a Daubechies filter with an order 5. Upon the analysis of the phase A voltage, the filters that produced the optimum features were *db5, db6, db7, db8* and their corresponding resolution levels were 6, 7, 6 and 6 respectively.

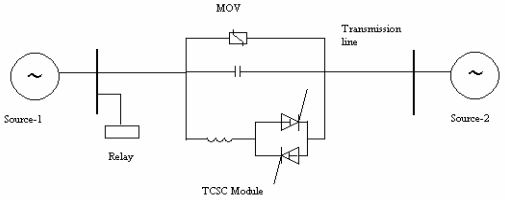

Fig. 4. The TCSC incorporated transmission line

Similarly, for the phase B voltage it was *db4* with resolution levels 5, 7 and 8 and *db7* with resolution level 5. For phase C it were *db4* and *db5* with resolution levels 7and 6 respectively. The same filters and their corresponding levels were also used for other faults and the results obtained were outstanding even thought the original analysis was for the a-g fault.

VI. Results

The SVM used is a two class SVM [13] for the identification of the faulty section. The output for the cases

TABLE I
PREDICTION RATES FOR INDIVIDUAL FAULT TYPES

Fault	Kernel	Cases tested by SVM	Value of C, g	Accuracy
a-g	RBF	1900	10,1	100%
b-g	RBF	1900	10,1	100%
c-g	RBF	1900	10,1	100%
a-b-g	RBF	1900	10,5	99.74%
a-c-g	RBF	1900	10,1	100%
b-c-g	RBF	1900	10,1	100%
a-b-c-g	RBF	1900	10,5	98.1%
a-b	RBF	1900	10,5	99.84%
a-c	RBF	1900	10,1	100%
b-c	RBF	1900	10,1	100%

that are before the TCSC is set to '1' and '-1' for the cases after the TCSC module. The inputs to the SVM are the maximum values of the outputs obtained after the feature extraction process for the three phase voltages. 50% of the data generated was used for training and the rest for testing. Table I shows the results after testing the SVM on unknown data as well as the values of the parameters C and g of the RBF kernel which was used to train the SVM. A total of 19000 cases were tested and a high accuracy was obtained. Before testing, a three-fold cross validation was also done for all the faults and the results were more or less the same as shown in the Table I. Only for the a-b-g fault was the accuracy a little low.

VII. CONCLUSION

From the Table I, it can be seen that the accuracy of the classifier in determining the faulty section is very high and is better than ones presented in [4] and [5]. Hence, the proposed method can be surely used for fault zone detection successfully.

APPENDIX

Generator Details:

Base kVA: 100000; Base kV: 400
Base frequency: 50 Hz.

Transmission line details:

Positive sequence impedance, Z(1) = 9.78+j110.23 Ω.
Zero sequence impedance, Z(0) = 96.45+j335.26 Ω.

REFERENCES

[1] Y. H. Song, A. T. Johns, Q. Y. Xuan, "Artificial neural-network based protection scheme for controllable series-compensated EHV transmission line", *Proc. Inst. Elect. Eng.-Gener. Transm. Distrib.*, vol. 143, no. 6, pp. 535-540, Nov. 1996.

[2] B. Kasztenny,"Distance protection of series compensated lines' problems and solutions,"*Proc. 28th Annu. Western Protective Relay Conf.*, Spokane, WA, Oct. 22-25,2001,pp.1-36.

[3] D. Nvosel, A. Phadke, M. M. Saha, and S. Lindhal, "Problems and solutions for microprocessor protection of series compensated lines," *Proc. 6th Inf. Conf. Developments in Power System Protection*, Mar. 25-27, 1997, pp. 18-23, Conf Pub No. 434.

[4] D. W. P. Thomas and C. Christopulos, "Ultra-high speed protection of series compensated lines," *IEEE Trans. Power Del.*, vol. 7, no. 1, pp. 139-145, Jan. 1992.

[5] P.K. Dash, S.R. Samantray, and Ganapati Panda, "Fault Classification and Section Identification of an Advanced Series-Compensated Transmission Line Using Support Vector Machine", *IEEE Trans. on Power Del*., vol. 22, no. 1, pp. 67-73, Jan 2007.

[6] Urmil B. Parikh, Biswarup Das, Rudra Pratap Maheshwari,"Comnined Wavelet-SVM Technique for Fault Zone Detection in a Series Compensated Transmission Line," *IEEE Trans. Power Del.*, vol. 23, no. 4, Oct 2008.

[7] S. R. Samantaray ," Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line,"*IET Gener. Transm. Distrib.*,vol. 3, Iss.5, pp. 425-436, 2009.

[8] Vladimir N. Vapnik, "An Overview of Statistical Learning Theory", *IEEE Transactions on Neural Networks,* vol. 10, no.5, pp 988-999 , Sept. 1999

[9] C. Cortes and V. Vapnik, "Support Vector Networks", *Int. Proceedings of Machine Learning,* vol. 20, no. 3, pp 273-297, 1995

[10] V. Vapnik, *The Nature of Statistical Learning Theory,* Springer, 1995.

[11] C. C. Burges, "A tutorial on support vector machines for pattern recognition", *In Proceedings of Int. Conference on Data Mining and Knowledge Discovery,* vol. 2, no. 2, pp. 121-167, 1998.

[12] "*PSCAD/EMTDC Power Systems Simulation Manual*", 1997, Winnipeg, MB, Canada.

[13] Chang C. C. and Chin J. L., *LIBSVM:A library for Support Vector Machines,* 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Renju Gangadharan is a postgraduate student at the Department of Electrical Engineering, Indian Institute of Technology-Roorkee, India. He did his undergraduate studies at the Department of Electrical and Electronics, M.A. College of Engineering, Kothamangalam, India. He became a member of the IEEE in 2006. His research interests are Digital Signal Processing and pattern classification, Power System Operation and Control, and Machine Learning.

G. N. Pillai is an Associate Professor at the Department of Electrical Engineering, Indian Institute of Technology-Roorkee, India. He completed his Ph. D. from the Indian Institute of Technology-Kanpur, India in the year 2001. He worked as Research Officer at the University of Ulster, Newtownabbey, UK and has many publications in reputed international journals/conferences. His main research interests are in the area of Soft Computing Applications, Power System Operation and Control.

Indra Gupta is working as an Associate Professor at the Department of Electrical Engineering, Indian Institute of Technology-Roorkee, India. She completed her Ph. D. from the Indian Institute of Technology-Roorkee in 1996. She has many publications in reputed journal so far. Her research areas include Power System, Simulation, Process Control Applications, Microprocessor Applications, ANN and Online Control Applications.