CLASSIFICATION OF POWER NETWORK DISTURBANCES BY USING THE SUPPORT VECTOR MACHINE METHOD

Peter G.V. Axelberg  Irene Y.H Gu  Math H.J. Bollen  
1) University College of Borås,  Chalmers University of 1) STRI AB, 
501 90 Borås, Sweden  Technology, 771 80 Ludvika, Sweden 
2) Unipower AB  412 96 Göteborg, Sweden 2) Luleå University of 
P.O. Box 411, 441 28 Alingsås,  Technology, 931 87 Skellefteå 
Sweden  
peter.axelberg@hb.se  irenegu@s2.chalmers.se  math.bollen@stri.se

ABSTRACT

This paper proposes a new method for classifying five common types of power system disturbances by using Support Vector Machines (SVMs). Equipped with the statistical learning theory, a SVM is found to be attractive in statistical-based characterization and classification of power system disturbance data. Apart from descriptions and discussions of the above issues, the paper also includes experimental results by using the proposed SVMs for automatic classification of several types of disturbances. These results were obtained by using more than 1200 measured disturbance data recordings.

I. INTRODUCTION

Over the past two decades utilities worldwide have gone through radical changes. One big change is the deregulation of the energy market that has taken place in a number of countries worldwide. Another change is that today’s customers are more demanding than customers in the past. These changes have forced the utilities to become even more customer oriented where high network reliability and high power quality has been increasingly important to keep customers satisfied. As a direct consequence of this, to record and analyze voltage disturbances and other power quality phenomena has become an important issue in the design and operation of power systems. The aim of such measurements is not just to quantify the power quality but also to better understand the behavior of the power network. This in turn enables a further improvement of power quality and reliability. Disturbance data and power quality data has therefore become important information both for statistical purposes and as a decision tool in power-quality mitigation projects. Reliable disturbance and power quality information opens also up for pro-active maintenance approach with focus to increase the power network reliability. Currently most of the disturbance data is analyzed manually by power-quality specialists, by power-system operators or (in far too many cases) simply stored and not further processed. Whereas the need for analysis is generally accepted, a lot of time could be saved if disturbances that do not need special attention from a specialist could be classified automatically. Thereby, the specialists could focus on solving more sophisticated disturbance problems without loosing information. Thus, a robust method for automatic classification of disturbances is of highly demand.

A number of papers on automatic classification of voltage disturbances have been published during the last several years. These can be roughly classified into two groups:
• Classification of event waveforms, with typical classes included: "voltage dips", "interruptions", "transients" and "distortion". This work has its importance for the development of classification tools but has limited practical value. The classification of waveforms is in practice based on rather simple rules like the 10% residual voltage limit to distinguish between a voltage dip and an interruption.

• Classification of events based on their origin, with typical classes including "faults", "transformer energizing" and "capacitor energizing".

The classification methods used and under development can roughly be classified into the following three groups:

• Visual inspection by a power-quality expert. This is the method most commonly used in practice for classification after origin of a disturbance.

• Rule based systems, often implemented as an expert system, to distinguish between different classes of event. The classification of waveforms by a power-quality monitor is often based on a set of rules, albeit rather simple rules. The classification is in that case not only based on waveform characteristics but also on the triggering. An expert system for classification of voltage dips and interruptions based on their origin is presented in [1]. This work is more applicable and actually highly needed. However, the practical implementation is much more difficult and further aggravated by the lack of the large amounts of data needed.

• Statistical-based methods using advanced signal-processing techniques. The artificial neural network is by far the most popular method in literature often combined with a set of wavelet filters for feature extraction and fuzzy logic for the decision making. Such a method is used in [2] and [3] for a classification based on waveform and in [4] to distinguish between transformer energizing and transformer energizing. An overview of statistical classification methods is given in [8], where the support-vector machine (SVM) is recommended as to be further studied for power-quality disturbance classification.

This work is therefore based on the Support Vector Machine (SVM) algorithm. The algorithm has its origin in statistical learning theory and has some attractive theoretical and practical characteristics [5] - [7] (an introduction to the SVM algorithm is given in Section II). A number of articles have been published recently showing that SVM classifiers are efficient for classification purposes in various applications. However, in earlier works the training data and test data originated from the same population. If the SVM classification or any other classification method should be implementable in a commercial system it must be possible to obtain settings that are to a large extent system independent. For statistical classification methods this implies that the method should be pre-trained from factory with a global setting. The aim of the work presented in this paper has therefore been to verify if such a global setting is possible. For this a large number of recordings were obtained from two medium-voltage distribution systems in two different European countries. Data from one country was used to train the classifier, which next was tested on data from the other country.

This paper is organized as follows. Section II describes the basic theory of SVMs. Section III discusses the characterization of voltage disturbances and feature extraction used for efficient training of the SVM. Section IV introduces the proposed SVM classifier. Section V describes conducted experiments and results. Finally Section V gives the conclusion.
II. FUNDAMENTALS OF SUPPORT VECTOR MACHINES

The main purpose of the (binary) SVM algorithm used for classification is to construct an optimal decision function, \( f(x) \), that accurately predicts unseen data into two classes and minimizes the classification error:

\[
f(x) = \text{sign}(g(x))
\]

(1)

This is achieved by following the method of structural risk minimization (SRM) which states that the expected classification error \( R \) of unseen data is bounded by the sum of a training error rate (first term in (2)) and a term that depends on the Vapnik-Chervonenkis (VC) dimension \( h \) (second term in (2) [5]):

\[
R < \frac{t}{N} + \sqrt{\frac{\ln(2N/h)+1-\ln(\eta/4)}{N}}
\]

(2)

\( t \) is the number of training errors, \( N \) is the number of training samples and \( \eta \) is a confidence measure.

In the case of separable data the first term in (2) is zero and the second one is minimized resulting in good generalization performance of the SVM (i.e. good classification performance of unseen data). The function \( g(x) \) in (1) is the decision boundary and is derived from a set of training samples:

\[
X = \{x_1, x_2, \ldots, x_n\}, \ x \in \mathbb{R}^M
\]

(3)

where each training sample \( x_i \) has \( M \) features describing a particular pattern and belongs to one of the two classes

\[
Y = \{y_1, y_2, \ldots, y_n\}, \ y \in \{-1, 1\}
\]

(4)

The decision boundary is a hyperplane

\[
g(x) = \langle w, x \rangle + b
\]

(5)

where \( w \) and \( b \) shall be derived in such a way that unseen data is classified correctly. This is achieved by maximizing the margin of separation between the two classes. According to [5] this can be formulated as a Quadratic Programming (QP) optimization problem (6)

\[
\Phi(w, \xi) = \min \left\{ \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i \right\}
\]

subject to the constraint that all training samples are correctly classified (i.e. all training samples are placed on the margin or outside the margin), that is

\[
y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \quad i = 1, \ldots, n
\]

(7)

where \( \xi_i, i = 1, \ldots, n \) are non-negative slack variables. By minimizing the first term of (6) the complexity of the SVM is reduced, and by minimizing the second term the number of training errors is decreased. Parameter \( C \) in (6) is a regularization parameter and is pre-selected to be
the trade-off between the two terms in (6). The constrained QP problem defined in (6) and (7) is solved by introducing Lagrange multipliers $\alpha_i \geq 0$ and $\beta_i \geq 0$ and the Lagrange functional

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left[ y_i \left[ \langle w, x_i \rangle + b \right] - 1 + \xi_i \right] - \sum_{i=1}^{n} \beta_i \xi_i$$

(8)

According to the theory of QP optimization, it is better to solve this optimization problem by introducing the dual formulation of the problem as follows.

$$\max_{\alpha, \beta} W(\alpha, \beta) = \max_{\alpha, \beta} \left\{ \min_{w, b, \xi} L(w, b, \xi, \alpha, \beta) \right\}$$

(9)

where $\alpha_i$ and $\beta_i$ are the Lagrange multipliers. That means, the optimal solution is given by first minimizing $w$, $b$ and $\xi$ and thereafter maximizing with respect to $\alpha_i \geq 0$ and $\beta_i \geq 0$. By substituting (8) into (9), the problem can be transformed to its dual formulation given by:

$$\max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \right\}$$

(10)

and shall be maximized under the constraints,

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C \quad \text{for} \quad i = 1, \ldots, n$$

(11)

Furthermore, the vector $w$ has an expansion in terms of a subset of the training samples where the Lagrange multipliers $\alpha_i$ are non-zero. Those training samples will also meet the Karush-Kuhn-Tucker condition:

$$\alpha_i \left[ y_i \left[ \langle w, x_i \rangle + b \right] - 1 + \xi_i \right] = 0, \quad i = 1, \ldots, n$$

(12)

Expression (12) states that only the training vectors corresponding to non-zero Lagrange multipliers, the support vectors (SV), are needed to describe the hyperplane. In the case of linearly separable data all Support Vectors (SVs) will lie on the margin and hence the number of SVs can be very small. Consequently, the optimal hyperplane is determined by using only a subset of the training samples and the rest of the training samples are not needed. Consequently, the decision boundary, $g(x)$, is determined by using only a subset of the training samples and the rest of the training samples are not needed:

$$g(x) = \sum_{i=1}^{N} \alpha_i y_i \langle x, x_i \rangle + b$$

(13)

In the case where a linear boundary is inappropriate the SVM can map the input vector, $x$, to a higher dimensional feature space [5], [6]. This is achieved by introducing a kernel function $K(.,.)$ and obtain the following substitution in (10):

$$\langle x_i, x_j \rangle \rightarrow K(x_i, x_j)$$

(14)
This yield

\[
\max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\}
\]

(15)

and (15) shall be maximized under the constraints in (7) and the solution is provided by using a software package for solving optimization problems. The decision boundary \( g(x) \) in (13) is then modified by substituting \( \langle x, x_i \rangle \) with \( K(x, x_i) \):

\[
g(x) = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b
\]

(16)

Functions that satisfy Mercer’s theorem are used as kernel functions. Soft decision SVM is then applied. Examples of such kernels are given in Table I.

<table>
<thead>
<tr>
<th>Kernel type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial</td>
<td>( K(x, y) = (x \cdot y + \delta)^d )</td>
</tr>
<tr>
<td>Radial basis function (RBF)</td>
<td>( K(x, y) = e^{-\gamma |x-y|^2} )</td>
</tr>
<tr>
<td>Sigmoidal</td>
<td>( K(x, y) = \tanh(\kappa (x \cdot y) - \delta) )</td>
</tr>
</tbody>
</table>

Table I. Examples of Kernel functions.

The kernel parameters \( d, \gamma, \kappa \) and \( \delta \) are (like the regularization parameter \( C \)) input parameters to the SVM training process and shall be tuned to achieve sufficient classification performance.

III. CHARACTERIZATION AND FEATURE EXTRACTION

The voltage disturbances studied here are sudden changes in the voltage waveform caused by short circuits, overloads or starting of heavy loads like motors etc [9]. Fig 1 represents a typical voltage disturbance with its waveform representation (top figure) and its rms signature (bottom figure).

![Voltage disturbance with its waveform (top) and rms signature (bottom). Shadowed areas in the bottom diagram indicate the two transition segments.](image-url)
Different types of voltage disturbances give different patterns in the time- and frequency domain of the disturbance. This implies that automatic classification of voltage disturbances can be achieved if robust features that describe disturbances can be identified.

A voltage or current event recording can be divided into a number of transition and event segments [1, 8 Chapter 7]. Transition segments correspond to large sudden changes in signal amplitude and/or waveform. Event segments are the periods in between transition segments.

The rms signature shown in Fig. 1 can be divided into five segments: Segment I is the pre-disturbance segment which occurs before the triggering of the disturbance. Segment II is the first transition segment, which occurs just after the triggering has occurred. The voltage during segment III is stationary; it is followed by a second transition segment (segment IV) at the end of the disturbance. Finally segment V starts after the voltage has retained to about its pre-disturbance level. In segment II and segment IV the voltage signal is non-stationary and hence no information can be obtained for feature extraction. Also segment I and segment V are of limited interest in terms of feature extraction since the disturbance has not yet started or it has passed. Hence the remaining segment used for feature extraction is segment III where the voltage disturbance is in its most stationary phase. This segment normally contains information that is unique enough to distinguish between different types of disturbances. Furthermore, features can be extracted both from the waveform and from the rms representation. Examples of features used for classification purposes are: the minimum rms voltage; the rms voltage at selected time instants; harmonic components (magnitude and phase) at selected time instants; total harmonic distortion; symmetric components and the duration of the disturbance [8].

In this work we concentrate on extracting features from the most common types of voltage disturbances, which are described below.

**Voltage disturbance due to a fault**

A voltage disturbance due to a fault is characterized by a sharp drop of the voltage rms at the starting point of the disturbance and also a sharp recovery when the disturbance is cleared. In between these two sharp changes, the rms voltage nearly remains constant. In a three-phase system there are mainly four different types of voltage disturbances due to a fault. Typical rms patterns for these types are shown in Fig. 2. The disturbances due to a fault have i) voltage magnitude drops in one phase (Fig 2a) ii) voltage magnitude drops in two phases (Fig 2b), iii) voltage magnitude drops in three phases (Fig 2c) iv) voltage magnitude drops in two phases but one of these is more affected than the other one (Fig 2d). In the remaining of this paper these four disturbance types are referred to as disturbance type D1 to D4 respectively. Type D3 is due to three-phase faults, whereas types D1, D2 and D4 may be due to single-phase faults or phase-to-phase faults. Classes D1 and D2 are dips with a small characteristic phase angle jump; the characteristics phase-angle jump is high for class D4 (a more detailed discussion regarding the origin of voltage disturbances is given in [9]). In general, all four types of disturbances due to a fault show considerable changes in the rms signatures whilst the harmonics magnitudes and THD are low.
Fig. 2. Rms signatures of voltage disturbances due to a fault. Fig. 2a shows a voltage drop in one phase (D1), Fig 2b shows a voltage drop in two phases (D2), Fig 2c shows a voltage drop in three phases (D3) and Fig 2d shows a drop in two phases where one of the phases is more affected than the other phase (D4).

Voltage disturbance due to transformer energizing.

Another common type of voltage disturbance is due to transformer saturation [1,8,9]. This type of disturbance is shown in Fig. 3 and is characterized by a sharp drop followed by gradual recovery of the voltage rms to about its pre-disturbance level. In the remaining parts of this paper this type is referred to as disturbance type D5.

Fig. 3. Rms signature of a voltage disturbance due to transformer energizing (disturbance type D5).
Relatively high magnitudes of even harmonics (especially 2nd and 4th harmonics) are produced as the transformer enters and exits the saturation and this observation will be used as a feature.

No load switching events were recorded in either study as this concerned medium-voltage networks for public distribution.

Feature extraction

From the above observations features can be defined to be used as training inputs to the SVM classifier. These features are extracted from both the rms signature and from the frequency magnitude spectrum of the disturbances.

Each disturbance type has the following features:

- 20 rms values per phase equally spread in time over the event segment.
- The second, fifth and ninth harmonic voltage and the THD calculated in the middle of segment III.

This results in a feature vector of 72 components per disturbance class (i.e. 24 features per phase).

IV. THE SVM CLASSIFIER

The main purpose of the work presented in this paper was to investigate if a SVM classifier system is able to accurately classify data from one power network when the classifier system is trained on data originating from another power network. Therefore, recorded disturbance data were used from two different medium-voltage networks in two different countries in Europe (see Table II).

<table>
<thead>
<tr>
<th>Disturbance type (class)</th>
<th>Number of available disturbances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Network in country A</td>
</tr>
<tr>
<td>D1</td>
<td>141</td>
</tr>
<tr>
<td>D2</td>
<td>181</td>
</tr>
<tr>
<td>D3</td>
<td>251</td>
</tr>
<tr>
<td>D4</td>
<td>127</td>
</tr>
<tr>
<td>D5</td>
<td>214</td>
</tr>
</tbody>
</table>

*Table II. Number of disturbance data available for training and test of the SVM classifier.*

In order to classify the disturbance types D1-D5 five individual binary SVMs were trained to classify each specific disturbance type. To train the SVM classifier, the disturbance data from the power network in country A were split into two equal parts. Half of the data were used to train the SVMs and the other half were used as testing data in experiment 1. The training of the five SVMs was obtained in a 3-folded cross-validation process and the classifiers were using the RBF-kernel function. Table III shows the result of the training process in terms of selected parameters (regularization parameter C and kernel parameter γ) and the number of support vectors.
Finally, the trained SVMs were connected as a multilevel binary decision tree according to Fig. 4. SVM$_1$ is trained to classify class $D1$, SVM$_2$ is trained to classify class $D2$ but class $D1$ is removed from the training process since this class was already classified by SVM$_1$ one level up in the binary tree etc. This type of training scheme forms a SVM classifier that is referred to as a mutual exclusive classifier.

![Fig. 4. Binary decision tree based on individual binary SVM$_k$ classifiers for multiple class classification.](image)

V. EXPERIMENTS AND RESULTS

Two experiments were conducted to evaluate the performance of the proposed SVM classifier system. Experiment I was to ensure that the classifier system worked properly. This was confirmed by making a classification when training data and test data originated from the same source (i.e. from power network in country A). The second experiment was the important one since training data and test data originated from different power networks. Table IV shows the number of test data used in the two experiments.

<table>
<thead>
<tr>
<th>Disturbance type</th>
<th># disturbances Originating from power network in country A</th>
<th># disturbances Originating from power network in country B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D1$</td>
<td>71</td>
<td>471</td>
</tr>
<tr>
<td>$D2$</td>
<td>91</td>
<td>125</td>
</tr>
<tr>
<td>$D3$</td>
<td>126</td>
<td>14</td>
</tr>
<tr>
<td>$D4$</td>
<td>64</td>
<td>196</td>
</tr>
<tr>
<td>$D5$</td>
<td>108</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table IV. Number of disturbances used as test data.*
Experiment 1: Training data and test data originate from the same power network.

The result from the first experiment is given in the matrix shown in Fig. 5. Each row in the matrix is a disturbance type (e.g. the first row is the D1 disturbance etc.) and the columns give the result in terms of number of classifications per disturbance type. The number of correct classifications are thus in the diagonal elements. The column NC (Not Classified) indicates that the classifier failed to classify to any appropriate class.

<table>
<thead>
<tr>
<th>Disturbance type</th>
<th># classifications per disturbance type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
</tr>
<tr>
<td>D1</td>
<td>63</td>
</tr>
<tr>
<td>D2</td>
<td>0</td>
</tr>
<tr>
<td>D3</td>
<td>5</td>
</tr>
<tr>
<td>D4</td>
<td>0</td>
</tr>
<tr>
<td>D5</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig 5. Classification results when training data and test data originate from power network in country A. The matrix contains the classification results in terms of number of correct classifications (diagonal elements) and misclassifications (the numbers outside the diagonal elements).

Experiment 2: Training data and test data originate from different power networks.

This experiment determines if a global setting is achievable. Compared to the previous experiment, a larger number of test data were available for disturbance types D1, D2 and D4. However, for type D3 only 14 disturbances were available and for type D5 no disturbances at all were available. The result of the conducted test is given in the matrix in Fig. 6.

<table>
<thead>
<tr>
<th>Disturbance type</th>
<th># classifications in each disturbance type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
</tr>
<tr>
<td>D1</td>
<td>463</td>
</tr>
<tr>
<td>D2</td>
<td>0</td>
</tr>
<tr>
<td>D3</td>
<td>2</td>
</tr>
<tr>
<td>D4</td>
<td>0</td>
</tr>
<tr>
<td>D5</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig 6. Classification results when training data and test data originate from two different power networks.

Analysis of the results

Table V shows the classification accuracy of the two tests in terms of detection rate (i.e. the number of correct classified disturbances divided by the total number of disturbances) per disturbance type.

<table>
<thead>
<tr>
<th>Disturbance type (class)</th>
<th>Detection rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment 1</td>
</tr>
<tr>
<td>D1</td>
<td>88.7%</td>
</tr>
<tr>
<td>D2</td>
<td>92.3%</td>
</tr>
<tr>
<td>D3</td>
<td>89.7%</td>
</tr>
<tr>
<td>D4</td>
<td>98.4%</td>
</tr>
<tr>
<td>D5</td>
<td>96.3%</td>
</tr>
</tbody>
</table>

Table V. Detection rate results from the two tests.
We observe from the results given in Table V that the overall detection rate was around 90% for both experiments, which is sufficiently high. There is no significant difference in detection rate between test data from the same network and test data from another network. The differences between the two experiments are probably due to stochastic variations related to the small data sample. The low detection rate for class D4 in experiment 2 was due to only three (out of 14) misclassifications.

VI. CONCLUSION AND FURTHER WORK

This paper proposes a method based on statistical learning and SVMs for classification of five common types of voltage disturbances. The results from the conducted experiments showed high average classification rate from all our tests. Even more encouraging was that the proposed SVM classifier shows high performance also when training data and test data originate from different networks. This indicates that it is possible to develop global settings for a power-quality disturbance classifier based on SVM.

However, the accuracy of the classifier system is very much dependent on the quality of the features selected to be used by the SVMs classifier. Our test results have shown that using features from both time domain (RMS signatures) and frequency domain (harmonic magnitudes and total harmonic distortion) effectively characterizes the disturbance classes used in the tests. Finally, the proposed SVM classifier based on a binary decision tree structure has shown to be effective for multi-class classification. Such a classifier system is shown to be flexible in terms of adding new disturbance types to the classification system.

VII. ACKNOWLEDGEMENTS

Thanks are due to Jonny Carlsson, Unipower AB, Alastair Ferguson, Scottish Power Power Systems and Arne Berlin and Thomas Gustafsson, Vattenfall Eldistribution AB, Sweden, for their contributions.

VIII. REFERENCES