Research Article

Classification of Underlying Causes of Power Quality Disturbances: Deterministic versus Statistical Methods

Math H. J. Bollen,[1, 2] Irene Y. H. Gu,[3] Peter G. V. Axelberg,[3] and Emmanouil Styvaktakis[4]

[1] *STRI AB, 771 80 Ludvika, Sweden*
[2] *EMC-on-Site, Luleå University of Technology, 931 87 Skellefteå, Sweden*
[3] *Department of Signals and Systems, Chalmers University of Technology, 412 96 Gothenburg, Sweden*
[4] *The Hellenic Transmission System Operator, 17122 Athens, Greece*

This paper presents the two main types of classification methods for power quality disturbances based on underlying causes: deterministic classification, giving an expert system as an example, and statistical classification, with support vector machines (a novel method) as an example. An expert system is suitable when one has limited amount of data and sufficient power system expert knowledge; however, its application requires a set of threshold values. Statistical methods are suitable when large amount of data is available for training. Two important issues to guarantee the effectiveness of a classifier, data segmentation, and feature extraction are discussed. Segmentation of a sequence of data recording is preprocessing to partition the data into segments each representing a duration containing either an event or a transition between two events. Extraction of features is applied to each segment individually. Some useful features and their effectiveness are then discussed. Some experimental results are included for demonstrating the effectiveness of both systems. Finally, conclusions are given together with the discussion of some future research directions.

1. INTRODUCTION

With the increasing amount of measurement data from power quality monitors, it is desirable that analysis, characterization, classification, and compression can be performed automatically [1–3]. Further, it is desirable to find out the cause of each disturbance, for example, whether a voltage dip is caused by a fault or by some other system event such as motor starting or transformer energizing. Designing a robust classification for such an application requires interdisciplinary research, and requires efforts to bridge the gap between power engineering and signal processing. Motivated by the above, this paper describes two different types of automatic classification methods for power quality disturbances: expert systems and support vector machines.

There already exists a significant amount of literature on automatic classification of power quality disturbances, among others [4–24]. Many techniques have further been developed for extracting features and characterization of power quality disturbances. Feature extraction may apply directly to the original measurements (e.g., RMS values), from some transformed domain (e.g., Fourier and wavelet transforms,

and subband filters) or from the parameters of signal models (e.g., sinusoid models, damped sinusoid models, AR models). These features may be combined with neural networks, fuzzy logic, and other pattern recognition methods to yield classification results.

Among the proposed methods, only a few systems have shown to be significant in terms of being highly relevant and essential to the real world problems in power systems. For classification and characterizing power quality measurements, [6, 15, 16] proposed a classification using wavelet-based ANN, HMM, and fuzzy systems; [19, 20] proposed an event-based expert system by applying event/transient segmentation and rule-based classification of features for different events; and [13] proposed a fuzzy expert system. Each of these methods is shown to be suitable for one or several applications, and is promising in certain aspects in the applications.

1.1. Some general issues

Despite the variety of classification methods, two key issues are associated with the success of any classification system.
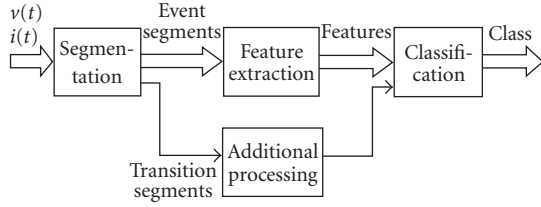
FIGURE 1: The processes of classification of power quality disturbances.

(i) Properly handling the data recordings so that each individual event (the term "event segment" will be used later) is associated with only one (class of) underlying cause. The situation were one event (segment) is due to a sequence of causes should be avoided.

(ii) Selecting suitable features that make the underlying causes effectively distinguishable from each other. It is counter-productive to use features that have the same range of values for all classes.

Extracting "good" features is strongly dependent on the available power system expertise, even for statistical classifiers. There exists no general approach on how the features should be chosen.

It is worth to notice a number of other issues that are easily forgotten in the design of a classifier. The first is that the goal of the classification system must be well formulated: is it aimed at classifying the type of voltage disturbance, or the underlying causes of the disturbances? It is a common mistake to mix the types of voltage disturbance (or phenomena) and their underlying causes. The former can be observed directly from the measurements, for example, interruption, dip, and standard classification methods often exist. While determining, the latter is a more difficult and challenging task (e.g., dip caused by fault, transformer energizing), and is more important for power system diagnostics. Finding the underlying causes of disturbances not only requires signal analysis, but often also requires the information of power network configuration or settings. Further, many proposed classification methods are verified by using simulations in a power-system model. It is important to notice that these models should be meaningful, consistent, and close to reality. Deviating from such a practice may make the work irrelevant to any practical application.

It is important to emphasize the integration of essential steps in each individual classification system as outlined in the block diagram of Figure 1.

Before the actual classification can take place, appropriate features have to be extracted as input to the classifier. Segmentation of the voltage and/or current recordings should take place first, after which features are mainly obtained from the event segments, with additional information from the processing of the transition segments.

1.2. Deterministic and statistical classifiers

This paper concentrates on describing the deterministic and statistical classification methods for power system disturbances. Two automatic classification methods for power quality disturbances are described: expert systems as a deterministic classification example, and support vector machines as a statistical classification example.

(i) Rule-based expert systems form a deterministic classification method. This method finds its application when there is a limited amount of data available, however, there exists good prior knowledge from human experts (e.g., from previously accumulated experience in data analysis), from which a set of rules can be created based on some previous expertise to make a decision on the origins of disturbances. The performance of the classification is much dependent on the expert rules and threshold settings. The system is simple and easy to implement. The disadvantage is that one needs to fine tune a set of threshold values. The method further leads to a binary decision. There is no probability on whether the decision is right or wrong or on the confidence of the decision.

(ii) Support vector machine classifiers are based on the statistical learning theory. The method is suitable for applications when there are large amounts of training data available. The advantages include, among others, that there are no thresholds to be determined. Further, there is a guaranteed upper bound for the generalization performance (i.e., the performance for the test set). The decision is made based on the learned statistics.

1.3. Structure of the paper

In Section 2, segmentation methods, including model residual and RMS sequence-based methods, are described. Through examples, Section 3 serves as the "bridge" that translates the physical problems and phenomena of power quality disturbances using power system knowledge into signal processing "language" where feature-based data characterization can then be used for distinguishing the underlying causes of disturbances. Section 4 describes a rule-based expert system for classification of voltage disturbances, as an example of deterministic classification systems. Next, Section 5 presents the statistical-based classification method using support vector machines along with a novel proposed method, which serves as an example of the statistical classification systems. Some conclusions are then given in Section 6.

2. SEGMENTATION OF VOLTAGE WAVEFORMS

For analyzing power quality disturbance recordings, it is essential *to partition the data into segments*. Segmentation, which is widely used in speech signal processing [25], is found to be very useful as a preprocessing step towards analyzing power quality disturbance data. The purpose of the segmentation is to divide a data sequence into stationary and nonstationary parts, so that each segment only belongs to one disturbance event (or one part of a disturbance event) which is caused by a single underlying reason. Depending

on whether the data within a segment is stationary, different signal processing strategies can then be applied. In [20], a typical recording of a fault-induced voltage dip is split into three segments: before, during, and after the fault. The divisions between the segments correspond to fault initiation and fault clearing. For a dip due to motor starting, the recording is split to only two segments: before and after the actual starting instant. The starting current (and thus the voltage drop) decays gradually and smoothly towards the new steady state.

Figure 2 shows several examples where the residuals of Kalman filter have been used for the segmentation. The segmentation procedure divides the voltage waveforms into parts with well-defined characteristics.

2.1. Segmentation based on residuals from the data model

One way to segment a given disturbance recording is to use the model residuals, for example, the residuals from a sinusoid model, or an AR model. The basic idea behind this method is that when a sudden disturbance appears, there will be a mismatch in the model, which leads to a large model error (or residual). Consider the harmonic model to describe the voltage waveform:

$$z(n) = \sum_{k=1}^{N} A_i \cos\left(2\pi n f_k + \phi_k\right) + v(n), \qquad (1)$$

where $f_k = 2\pi k f_0 / f_s$ is the kth harmonic frequency, f_0 is assumed to be the power system fundamental frequency, and f_s the sample frequency. Here, the model order N should be selected according to how many harmonics are required to accommodate as being nonsignificant disturbances.

To detect the transition points, the following *measure of change* is defined and used to extract the average Kalman filter residual $e(n) = z(n) - \hat{z}(n)$ within a short window of size w as follows:

$$d(n) = \left(\frac{1}{w} \sum_{i=n-w/2}^{n+w/2} \left(z(i) - \hat{z}(i)\right)\right)^2, \qquad (2)$$

where $\hat{z}(n)$ is the estimate $z(n)$ from the Kalman filter. If $d(n)$ is prominent, there is a mismatch between the signal and the model and a transition point is assumed to have been detected. Figure 3 shows an example where the transition points were extracted by utilizing the residuals of a Kalman filter of $N = 20$. This is a recording of a multistage voltage dip as measured in a medium voltage network.

The detection index $d(n)$ is obtained by using the residuals of three Kalman filters (one for each phase). Then the three detection indices are combined into one by considering at each time instant the largest of the three indices. In such a way, the recordings are split to *event segments* (where the detection index is low) and *transition segments* (where the detection index is high). The order of Kalman filters is selected as being able to accommodate the harmonics that are caused by events like transformer saturation or arcing, which was suggested as $N = 20$ in [20].

2.2. Segmentation based on time-dependent RMS sequences

In case only RMS voltage versus time is available, segmentation remains possible with these time-dependent RMS sequence as input. An RMS sequence is defined as

$$\mathcal{V}_{\mathrm{RMS}}(t_k) = \sqrt{\frac{1}{N} \sum_{t=t_{k-N+1}}^{t_k} v^2(t)}, \quad t_k = t_0, t_1, \ldots, \qquad (3)$$

where $v(t_k)$ is the voltage/current sample, and N is the size of the sliding window used for computing the RMS. Rms sequence-based segmentation uses a similar strategy as the method discussed above; however the *measure of change* is computed from the derivatives of RMS values instead of from model residuals. The segmentation can be described by the following steps.

(1) Downsampling an RMS sequence

Since the time-resolution of an RMS sequence is low and the difference between two consecutive RMS samples is relatively small, the RMS sequence is first downsampled before computing the derivatives. This will reduce both the sensitivity of the segmentation to the fluctuations in RMS derivatives and the computational cost. In general, an RMS sequence with a larger downsample rate will result in fewer false segments (or split of segments) however with a lower time resolution of segment boundaries. Conversely, for an RMS sequence with a smaller downsample rate, the opposite holds. Practically the downsample rate m is often chosen empirically, for example, for the examples in this subsection, $m \in [N/16, N]$ is chosen (N is the number of RMS samples in one cycle). In many cases only a limited number of RMS values per cycle are stored (two according to the standard method in IEC 61000-4-30) so that further downsampling is not needed. For notational convenience we denote the downsampled RMS sequence as

$$\mathcal{V}_{\mathrm{RMS}}(\widetilde{t}_k), \quad \widetilde{t}_k = \frac{t_k}{m}. \qquad (4)$$

(2) Computing the first-order derivatives

A straightforward way to detect the segmentation boundaries is from the changes of RMS values, for example, using the first-order derivative

$$\mathcal{M}_{\mathrm{RMS}}^{j}(\widetilde{t}_k) = \left| \mathcal{V}_{\mathrm{RMS}}^{j}(\widetilde{t}_k) - \mathcal{V}_{\mathrm{RMS}}^{j}(\widetilde{t}_{k-1}) \right|, \qquad (5)$$

where $j = a, b, c$ indicate the different phases. Consider either a single-phase or a three-phase measurement, the measure of changes $\mathcal{M}_{\mathrm{RMS}}$ in RMS values is defined by

$$\mathcal{M}_{\mathrm{RMS}}(\widetilde{t}_k)$$
$$= \begin{cases} \mathcal{M}_{\mathrm{RMS}}^{a}(\widetilde{t}_k) & \text{for 1 phase,} \\ \max\left(\left(\mathcal{M}_{\mathrm{RMS}}^{a}, \mathcal{M}_{\mathrm{RMS}}^{b}, \mathcal{M}_{\mathrm{RMS}}^{c}\right)(\widetilde{t}_k)\right) & \text{for 3 phases.} \end{cases}$$
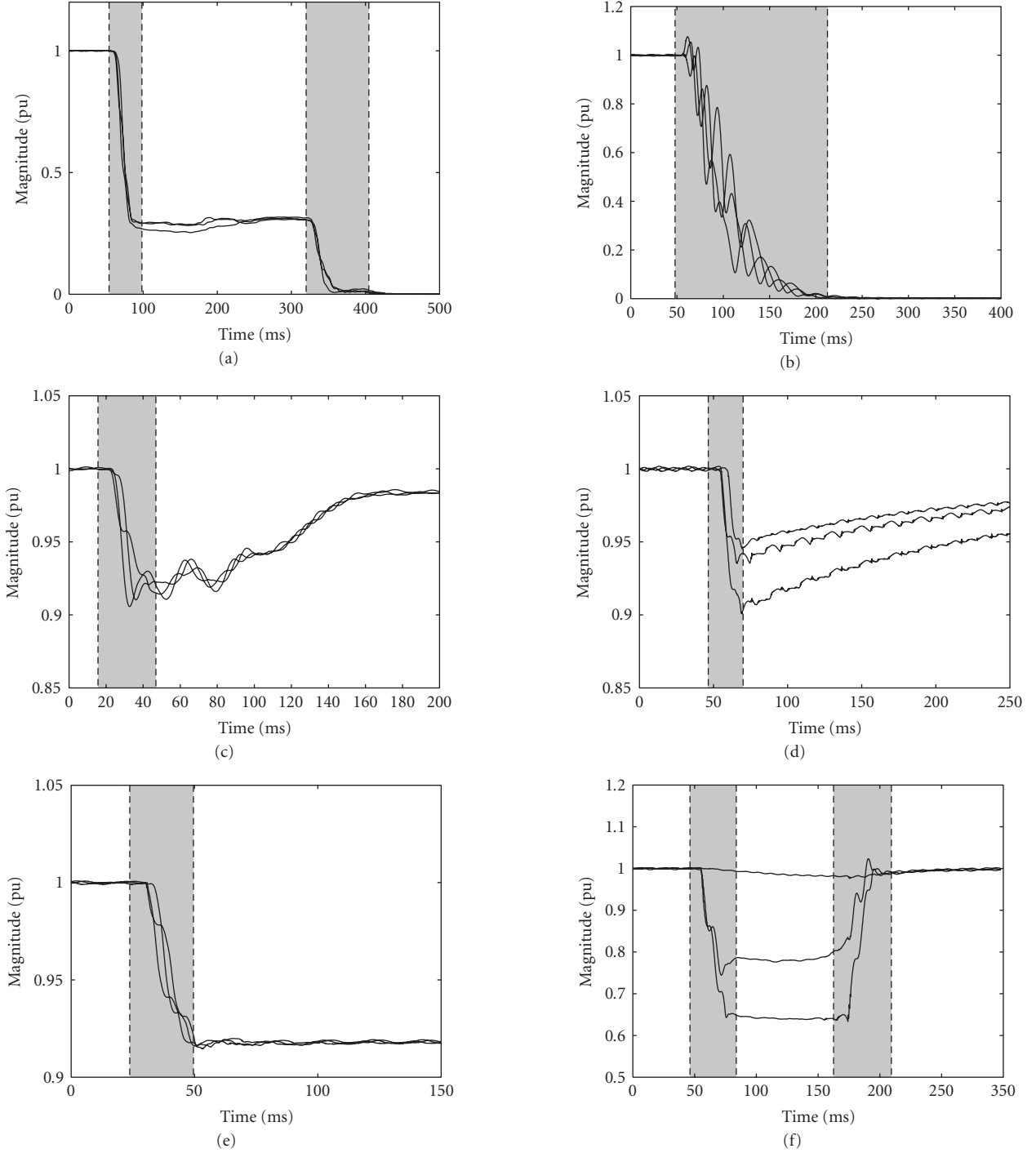$$\qquad (6)$$

FIGURE 2: RMS voltage values versus time (shadowed parts: transition segments). (a)–(f): (a) an interruption due to fault; (b) nonfault interruption; (c) induction motor starting; (d) transformer saturation; (e) step change; and (f) single stage voltage dip due to fault.

(3) Detecting the boundaries of segments

A simple step is used to detect the boundaries of segments under two hypotheses:

$$H_0(\text{event-segment}) : \mathcal{M}_{\text{RMS}}(\tilde{t}_k) < \delta,$$
$$H_1(\text{transition-segment}) : \mathcal{M}_{\text{RMS}}(\tilde{t}_k) \geq \delta, \qquad (7)$$

where δ is a threshold. A transition segment starts at the first $\tilde{t}_k$ for which H_1 is satisfied, and ends at the first $\tilde{t}_k$ for which $\mathcal{M}_{\text{RMS}}(\tilde{t}_k) < \delta$ occurs after a transition segment is detected.

It is recommended to use waveform data for feature extraction whenever available, as extracting features from the RMS voltages leads to loss of information. It is however possible, as shown in [20], to perform segmentation based on
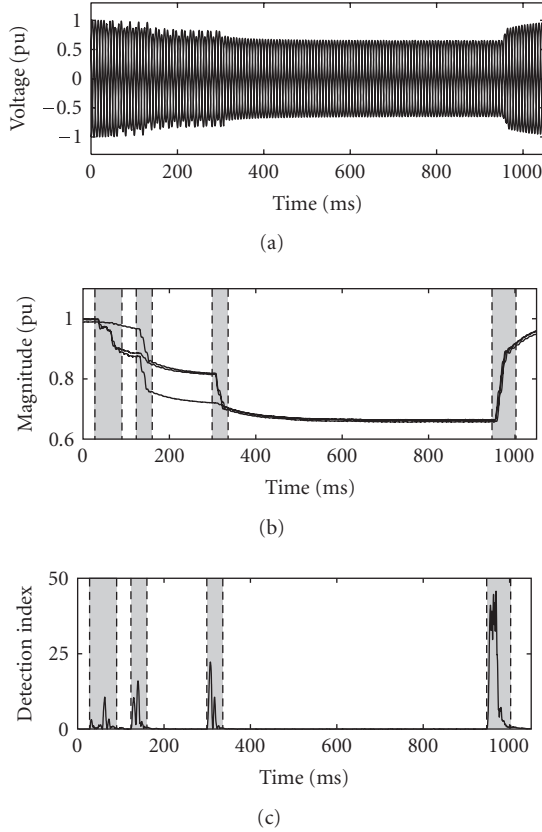
(a)

(b)

(c)

FIGURE 3: Using the Kalman filter to detect sudden changes ($N = 20$): (a) original voltage waveforms from 3 phases; (b) the detected transition points (marked on the fundamental voltages) that are used as the boundaries of segmented blocks; (c) the detection index by considering all three phases.

recorded RMS sequences. The performance of the resulting classifier is obviously less than that for a classifier based on the full waveform data.

3. UNDERSTANDING POWER QUALITY DISTURBANCES: UNDERLYING CAUSES AND THEIR CHARACTERIZATION

Characterizing the underlying causes of power quality disturbances and extracting relevant features based on the recorded voltages or currents is in general a difficult issue: it requires understanding the problems and phenomena of power quality disturbances using power system knowledge. A common and essential step for successfully applying signal processing techniques towards any particular type of signals is largely dependent on understanding the nature of that signal (e.g., speech, radar, medical signals) and then "translating" them into the problems from signal processing viewpoints. This section, through examples, contributes to understanding and "translating" several types of power quality disturbances into the perspective of signal processing. With visual inspection of the waveform or the spectra of disturbances, the success rate is very much dependent on a person's understand-
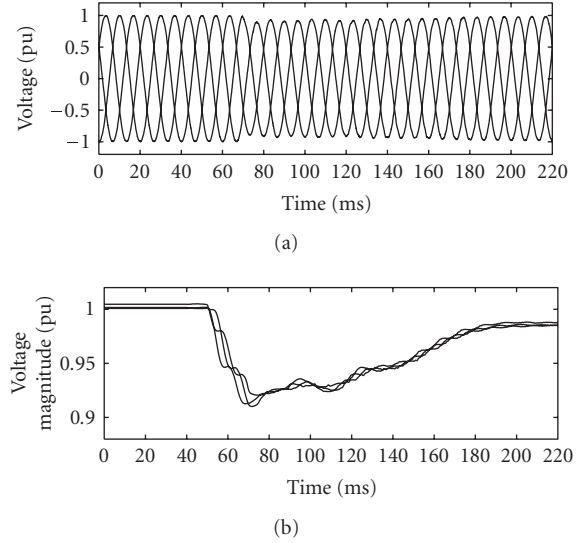

(a)

(b)

FIGURE 4: Induction motor starting: (a) voltage waveforms; (b) voltage magnitude (measurement in a 400 V network).

ing and previous knowledge of disturbances in power systems. An automatic classification system should be based at least in part on this human expert knowledge. The intension is to give some examples of voltage disturbances that are caused by different types of underlying reasons. One should be aware that this list is by far complete. It should further be noted that the RMS voltage as a function of time (or, RMS voltage shape) is used here to present the events, even though the features may be better extracted from the actual waveforms or from some other transform domain. We will emphasize that RMS sequences are by far the only time-dependent characteristics to describe the disturbances; many other characteristics can also be exploited [5].

Induction motor starting

The voltage waveform and RMS voltages for a dip due to induction motor starting are shown in Figure 4. A sharp voltage drop, corresponding to the energizing of the motor, is followed by gradual voltage recovery when the motor current decreases towards the normal operating current. As an induction motor takes the same current in the three phases, the voltage drop is the same in the three phases.

Transformer energizing

The energizing of a transformer gives a large current, related to the saturation of the core flux, which results in a voltage dip. An example is shown in Figure 5, where one can observe that there is a sharp voltage drop followed by gradual voltage recovery. As the saturation is different in the three phases, so is the current. The result is an unbalance voltage dip; that is, a dip with different voltage magnitude in the three phases. The dip is further associated with a high harmonic distortion, including even harmonics.
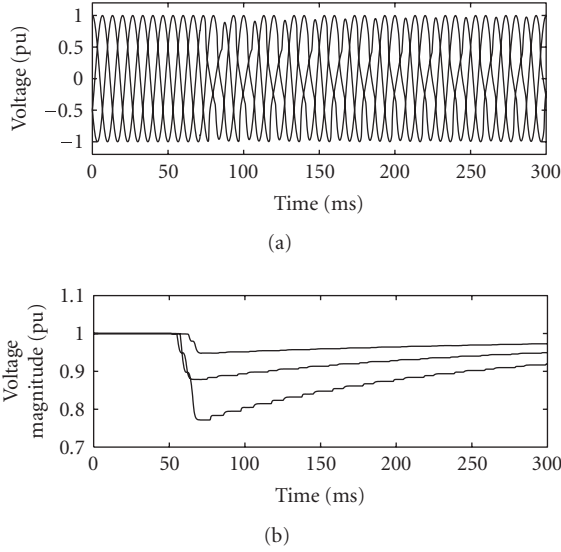
Figure 5: Voltage dip due to transformer energizing: (a) voltage waveforms; (b) voltage magnitude (EMTP simulation).
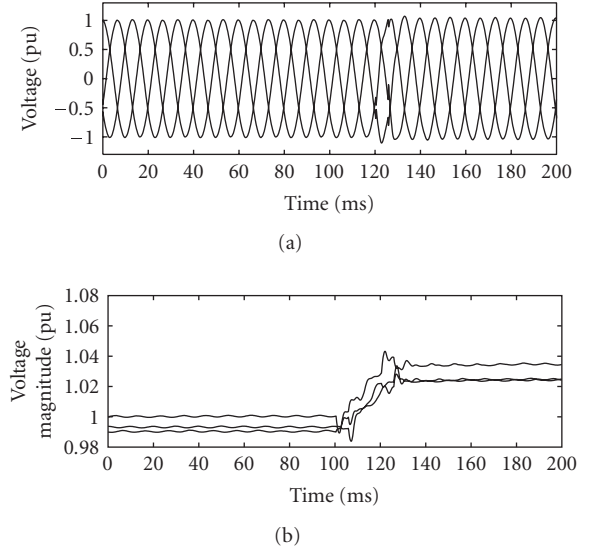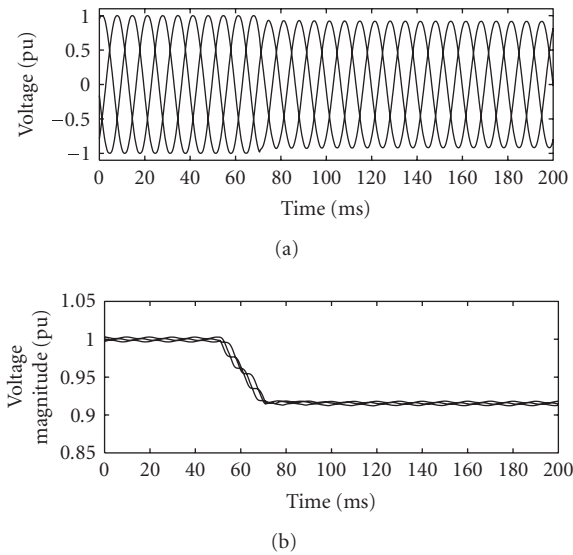

Figure 6: Voltage disturbance due to load switching: step change measurement: (a) voltage waveforms; (b) voltage magnitude (measurement in an 11 kV network).

Load switching

The switching of large loads gives a drop in voltage, but without the recovery after a few seconds, see Figure 6. The equal drop in the three phases indicates that this event was due to the switching of a three-phase load. The disconnection of the load will give a similar step in voltage, but in the other direction.

Capacitor energizing

Capacitor energizing gives a rise in voltage as shown in Figure 7, associated with a transient, in this example a minor

Figure 7: Voltage disturbance due to capacitor energizing: (a) voltage waveforms; (b) voltage magnitude (measurement in a 10 kV network).

transient, and often a change in harmonic spectrum. Capacitor banks in the public grid are always three phase so that the same voltage rise will be observed in the three phases. The recording shown here is due to synchronized capacitor energizing where the resulting transient is small. Nonsynchronized switching gives severe transients which can be used as a feature to identify this type of event. Several types of loads are equipped with a capacitor, for example, as part of their EMI-filter. The event due to switching of these loads will show similar characteristics to capacitor energizing. Capacitor de-energizing will give a drop in voltage in most cases without any noticeable transient.

Voltage dip due to a three-phase fault

The most common cause of severe voltage dips in distribution and transmission systems are symmetrical and nonsymmetrical faults. The large fault current gives a drop in voltage between fault initiation and the clearing of the fault by the protection. An example of a voltage dip due to a symmetrical (three-phase) fault is shown in Figure 8: there is a sharp drop in voltage (corresponding to fault initiation) followed by a period with constant voltage and a sharp recovery (corresponding to fault clearing). The change in voltage magnitude has a rectangular shape. Further, all three phases are affected in the same way for a three-phase fault.

Voltage dip due to an asymmetric fault

Figure 9 shows a voltage dip due to an asymmetric fault (a fault in which only one or two phases are involved). The dip in the individual phases is the same as for the three-phase fault, but the drop in voltage is different in the three phases.
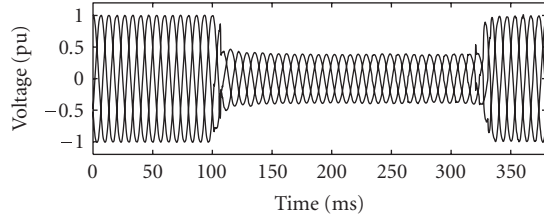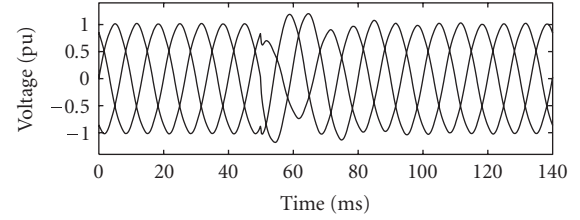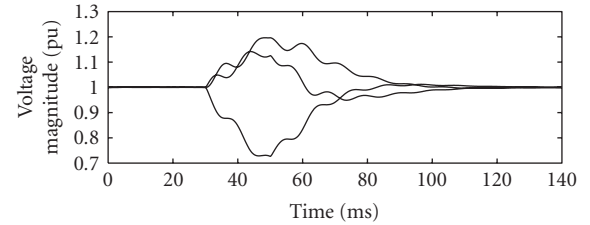
(a)
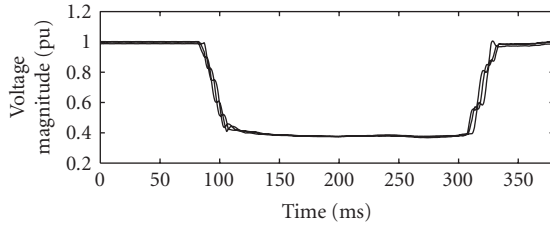
(b)

FIGURE 8: Voltage dip due to a symmetrical fault: (a) voltage waveforms; (b) voltage magnitude (measurement in an 11 kV network).
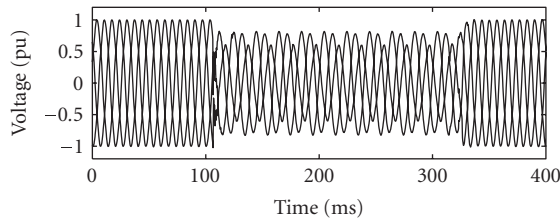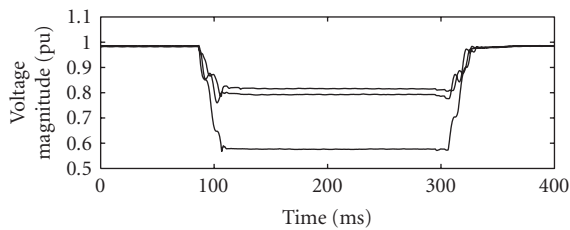

(a)

(b)

FIGURE 10: A self-extinguishing fault: (a) voltage waveforms; (b) voltage magnitude (measurement in a 10 kV network).

(a)

(b)

FIGURE 9: Voltage dip due to an asymmetrical fault: (a) voltage waveforms; (b) voltage magnitude (measurement in an 11 kV network).
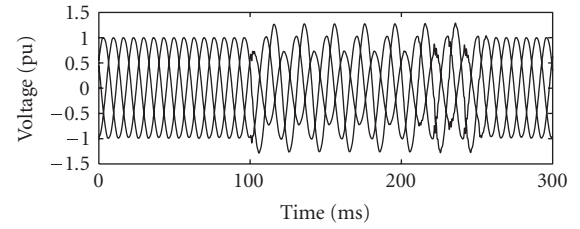
(a)
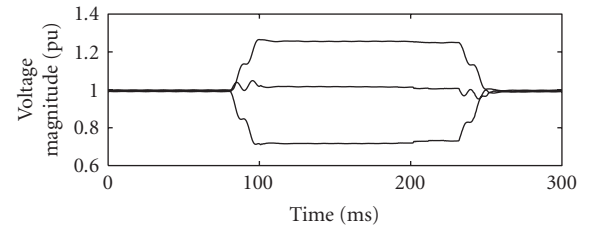
(b)

FIGURE 11: Overvoltage swell due to a fault: (a) voltage waveforms; (b) voltage magnitude (measurement in an 11 kV network).

Self-extinguishing fault

Figure 10 shows the waveform and RMS voltage due to a self-extinguishing fault in an impedance-earthed system. The fault extinguishes almost immediately, giving a low-frequency (about 50 Hz) oscillation in the zero-sequence voltage. This oscillation gives the overvoltage in two of the three-phase voltages. This event is an example where the customers are not affected, but information about its occurrence and cause is still of importance to the network operator.

Voltage swell due to earthfault

Earthfaults in nonsolidly-earthed systems result in overvoltages in two or three phases. An example of such an event is shown in Figure 11. In this case, the voltage rises in one phase, drops in another phase, and stays about the same in the third phase. This measurement was obtained in a low-resistance-earthed system where the fault current is a few times the nominal current. In systems with lower fault currents, typically both nonfaulted phases show an increase in voltage magnitude.

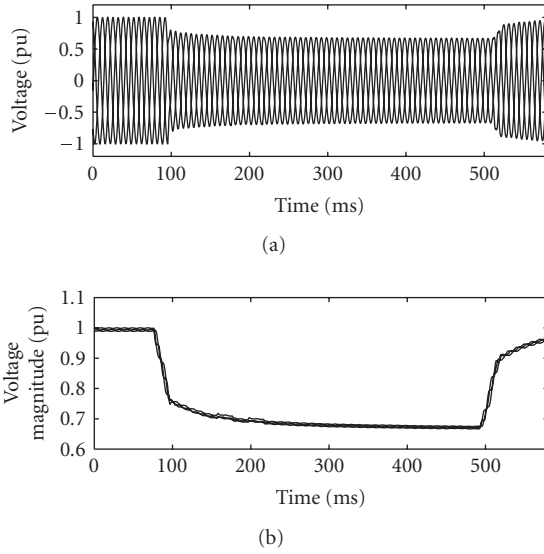FIGURE 12: Voltage dip due to fault, with the influence of induction motor load during a fault: (a) voltage waveforms; (b) voltage magnitude (measurement in an 11 kV network).

Induction motor influence during a fault

Induction motors are affected by the voltage drop due to a fault. The decrease in voltage leads to a drop in torque and a drop in speed which in turn gives an increase in current taken by the motor. In the voltage recording, this is visible as a slow decrease in the voltage magnitude. An example is shown in Figure 12: a three-phase fault with motor influence during and after the fault.

4. RULE-BASED SYSTEMS FOR CLASSIFICATION

4.1. Expert systems

A straightforward way to implement knowledge from power-quality experts in an automatic classification system is to develop a set of classification rules and implement these rules in an expert system. Such systems are proposed, for example, in [6, 7, 14, 19, 20].

This section is intended to describe expert systems through examples of some basic building blocks and rules upon which a sample system can be built. It is worth mentioning that the sample expert system described in this section is designed to only deal with a certain number of disturbance types rather than all types of disturbances. For example, arcing faults and harmonic/interharmonic disturbances are not included in this system, and will therefore be classified as unknown/rejected type by the system.

A typical rule-based expert system, shown in the block diagram of Figure 13, may consist of the following blocks.

(i) User interface

It is the interface where the data are fed as the input into a system (e.g., from the output of a power system monitor),
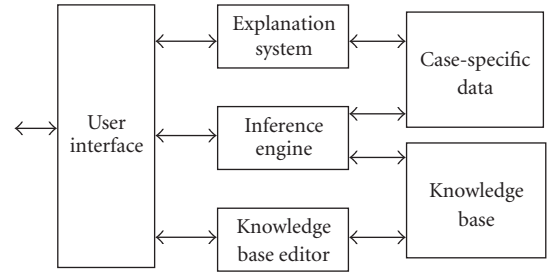
FIGURE 13: Block diagram of an expert system.

the classification or diagnostic results are the output through the interface (e.g., a computer terminal).

(ii) Inference engine

An inference engine performs the reasoning with the expert system knowledge (or, rules) and the data from a particular problem.

(iii) Explanation system

An explanation system allows the system to explain the reasoning to a user.

(iv) Knowledge-base editor

The system may sometimes include this block so as to allow a human expert to update or check the rules.

(v) Knowledge base

It contains all the rules, usually a set of IF-THEN rules.

(vi) Case-specific data

This block includes data provided by the user and can also include partial conclusions or additional information from the measurements.

4.2. Examples of rules

The heart of the expert system consists of a set of rules, where the "real intelligence" by human experts is translated into "artificial intelligence" for computers. Some example of rules, using RMS sequences as input, are given below. Most of these rules can be deducted from the description of the different events given in the previous section.

Rule 1 (interruption). IF at least two consecutive RMS voltages are less than 0.01 pu, THEN the event is an interruption.

Rule 2 (voltage swell). IF at least two consecutive RMS voltages are more than 1.10 pu, AND the RMS voltage drops below 1.10 pu within 1 minute, THEN the event is a voltage swell.

Rule 3 (sustained overvoltage). IF the RMS voltage remains above 1.06 pu for 1 minute or longer, AND the event is not a voltage swell, THEN the event is a sustained overvoltage.

Rule 4 (voltage dip). IF at least two consecutive RMS voltages are less than 0.90 pu, AND the RMS voltage rises above 0.90 pu within 1 minute, AND the event is not an interruption, THEN the event is a voltage dip.

Rule 5 (sustained undervoltage). IF the RMS voltage remains below 0.94 pu for 1 minute or longer, AND the event is not a voltage dip, AND the event is not an interruption, THEN the event is a sustained undervoltage.

Rule 6 (voltage step). IF the RMS voltage remains between 0.90 pu and 1.10 pu and the difference between two consecutive RMS voltages remains within 0.0025 pu of its average value for at least 20 seconds before and after the step, and the event is not a sustained overvoltage, AND the event is not a sustained undervoltage, THEN the event is a voltage step.

Note that these rules do allow for an event being both a voltage swell and a voltage dip. There are also events that possibly do not fall in any of the event classes. The inference engine should be such that both combined events and non-classifiable events are recognized. Alternatively, two additional event classes may be defined as "combined dip-swell" and "other events." For further classifying the underlying causes of a voltage dip, the following rules may be applied.

Rule 7 (voltage dip due to fault). IF the RMS sequence has a fast recovery after a dip (rectangular shaped recovery), THEN the dip is due to a fault (rectangular shaped is caused by protection operation).

Rule 8 (induction motor starting). IF the RMS sequences for all three phases of voltage recover gradually but with approximately the same voltage drop, THEN it is caused by induction motor starting.

Rule 9 (transformer saturation). IF the RMS sequences for all three phases of voltage recover gradually but with different voltage drop, THEN it is caused by transformer saturation.

4.3. Application of an expert system

A similar set of rules, but using waveform data as input, has been implemented in an expert system and applied to a large number of disturbances (however, limited to 9 types), obtained in a medium-voltage distribution system [19, 20].

The expert system is designed to classify each voltage disturbance into one of the 9 classes, according to its underlying cause. The list of underlying causes of disturbances being considered in this expert system includes

  (i) energizing,
 (ii) nonfault interruption,
(iii) fault interruption,
(iv) transformer saturation due to fault,
 (v) induction motor starting,
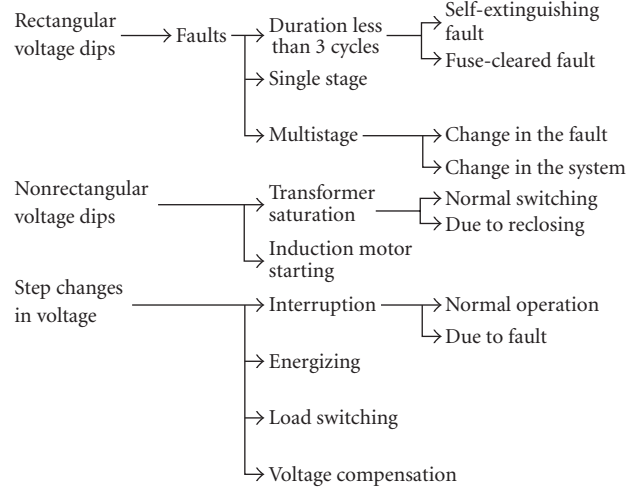
FIGURE 14: A tree structured inference process for classifying power system disturbance recordings.

 (vi) step change,
(vii) transformer saturation followed by protection operation,
(viii) single stage dip due to fault,
 (x) Multistage dip due to fault.

Some further analysis and classification are then applied, for example,

  (i) seven types of dip (type A, Ca, Cb, Cc, Da, Db, Dc, as defined in [3]);
 (ii) step change associated with voltage increase/decrease;
(iii) overvoltage associated with energizing/transformer saturation/step change/faults;
(iv) fault related to starting/voltage swell/clearing.

The tree-structured inference process for classifying the underlying causes is shown in Figure 14.

Table 1 shows the results from the expert system. Comparing with the ground truth (manually classified results from power system experts), the expert system has achieved approximately 97% of classification rate for a total of 962 disturbance recordings.

5. STATISTICAL LEARNING AND CLASSIFICATION USING SUPPORT VECTOR MACHINES

5.1. Motivations

A natural question arises before we describe SVM classifiers. Why should one be interested in an SVM when there are many other classification methods? Two main issues of interest in SVM classifiers are *the generalization performance* and *the complexity of classifier* which is a practical implementation concern.

When designing a classification system, it is natural that one would like the classifier to have a good *generalization performance* (i.e., the performance on the test set rather than on

TABLE 1: Classification results for 962 measurements.

Type of events	Number of classified events	
Energizing	104	
Interruption (by faults/nonfaults)	13	88
Transformer saturation (normal/protection operation)	119	6
Step change (increase/decrease)	15	21
Fault-dip (single stage)	455	
Fault-dip (multistage) (fault/system change)	56	56
Other nonclassified (short duration/overvoltage)	16	13

the training set). If one uses too many training samples, a classifier might be overfitted to the training samples. However, if one has too few training samples, one may not be able to obtain a sufficient statistical coverage to most possible situations. Both cases will lead to poor performance on the test data set. For an SVM classifier, there is a guarantee of the upper error bound on the test set based on statistical learning theory. Complexity of classifiers is a practical implementation issue. For example, a classifier, for example, a Bayesian classifier, may be elegant in theory, however, a high computational cost may hinder its practical use. For an SVM, the complexity of the classifier is associated with the so-called VC dimension.

An SVM classifier minimizes the generalization error on the test set under the structural risk minimization (SRM) principle.

5.2. SVMs and the generalization error

One special characteristic of an SVM is that instead of dimension reduction is commonly employed in pattern classification systems, the input space is nonlinearly mapped by $\Phi(\cdot)$ onto a high-dimensional feature space, where $\Phi(\cdot)$ is a kernel satisfying Mercer's condition. As a result of this, classes are more likely to be linearly separable in the high-dimensional space rather than in a low-dimensional space.

Let input training data and the output class labels be described as pairs $(\mathbf{x}_i, d_i)$, $i = 1, 2, \ldots, N$, $\mathbf{x}_i \in \mathbb{R}^{m_0}$ (i.e., m_0 dimensional input space) and $d_i \in Y$ (i.e., the decision space). As depicted in Figure 15, the spaces and the mappings for an SVM, a nonlinear mapping function $\Phi(\cdot)$ is first applied which maps the input-space $\mathbb{R}^{m_0}$ onto a high-dimensional feature-space $\mathcal{F}$,

$$\Phi : \mathbb{R}^{m_0} \longrightarrow \mathcal{F} \qquad \mathbf{x}_i \longmapsto \Phi(\mathbf{x}_i), \qquad (8)$$

where Φ is a nonlinear mapping function associated with a kernel function.
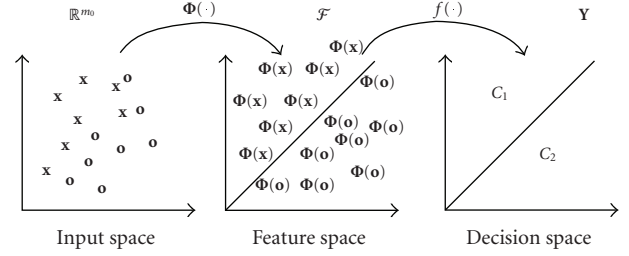

FIGURE 15: Different spaces and mappings in a support vector machine.

Then, another function $f(\cdot) \in \mathbb{F}$ is applied to map the high-dimensional feature space $\mathcal{F}$ onto a decision space,

$$f : \mathcal{F} \longrightarrow Y \qquad \Phi(\mathbf{x}_i) \longmapsto f(\Phi(\mathbf{x}_i)). \qquad (9)$$

The best function $f(\cdot) \in \mathbb{F}$ that may correctly classify a unseen example $(\mathbf{x}, d)$ from a test set is the one minimizing the expected error, or the generalization error,

$$R(f) = \int l[f(\Phi(\mathbf{x})), d] dP(\Phi(\mathbf{x}), d), \qquad (10)$$

where $l(\cdot)$ is the loss function, and $P(\Phi(\mathbf{x}), d)$ is the probability of $(\Phi(\mathbf{x}), d)$ which can be obtained if the probability of generating the input-output pair $(\mathbf{x}, d)$ is known. A loss (or, error) is occurred if $f(\Phi(\mathbf{x})) \neq d$.

Since $P(\mathbf{x}, d)$ is unknown, we cannot directly minimize (10). Instead, we try to estimate the function $f(\cdot)$ that is close to the optimal one from the function class $\mathbb{F}$ using the training set. It is worth noting that there exist many $f(\cdot)$ that give perfect classification on the training set; however they give different results on the test set.

According to VC theory [26–29], we choose a function $f(\cdot)$ that fits to the necessary and sufficient conditions for the consistency of empirical risk minimization,

$$\lim_{n \to \infty} P \left\{ \sup_f (R(f) - R_{\text{emp}}(f)) > \varepsilon \right\} = 0, \quad \forall \varepsilon > 0, \quad (11)$$

where the empirical risk (or the training error) is defined on the training set

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^{N} l(f(\Phi(\mathbf{x}_i)), d_i) \qquad (12)$$

and N is the total number of samples (or feature vectors) in the training set. A specific way to control the complexity of function class $\mathbb{F}$ is given by VC theory and the *structural risk minimization* (SRM) principle [27, 30]. Under the SRM principle, the function class $\mathbb{F}$ (and the function f) is chosen such that the upper bound of the generalization error in (10) is minimized. For all $\delta > 0$ and $f \in \mathbb{F}$, it follows that the bound of the generalization error

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{h(\ln(2N/h) + 1) - \ln(\delta/4)}{N}} \qquad (13)$$

holds with a probability of at least $(1 - \delta)$ for $N > h$, where h is the VC dimension for the function class $\mathbb{F}$. The VC dimension, roughly speaking, measures the maximum number of training samples that can be correctly separated by the class of functions $\mathbb{F}$. For example, N data samples can be labeled to 2^N possible ways in a binary class $Y = \{1, -1\}$ case, of which there exists at least one set of h samples that can be correctly classified to their class labels by the chosen function class $\mathbb{F}$.

As one can see from (13), minimization of $R(f)$ is obtained by yielding a small training error $R_{\mathrm{emp}}(f)$ (the 1st term) while keeping the function class as small as possible (the 2nd term). Hence, SVM learning can be viewed as seeking the best function $f(\cdot)$ from the possible function set $\mathbb{F}$ according to the SRM principle which gives the lowest upper bound in (13).

Choosing a nonlinear mapping function $\mathbf{\Phi}$ in (8) is associated with selecting a kernel function. Kernel functions in an SVM must satisfy Mercer's condition [28]. Roughly speaking, Mercer's condition states for which kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ there exists (or does not exist) a pair $\{\mathcal{F}, \mathbf{\Phi}\}$. Or, whether a kernel is a dot product in some feature space $\mathcal{F}$. Choosing different kernels leads to different types of SVMs. The significance of using kernel in SVMs is that, instead of solving the primal problem in the feature space, one can solve the dual problem in SVM learning which only requires the inner products of feature vectors rather than features themselves.

5.3. Soft-margin SVMs for linearly nonseparable classes

Consider a soft-margin SVM for linearly nonseparable classes. The strategy is to allow some training error $R_{\mathrm{emp}}(f)$ in the classifier in order to achieve less error on the test set, hence (13) is minimized. This is because zero training error can lead to overfitting and may not minimize the error on the test set.

An important concept in SVMs is the *margin*. A margin of a classifier is defined as the shortest distance between the separating boundary and an input training vector that can be correctly classified (e.g., classified as being $d_1 = +1$ and $d_2 = -1$). Roughly speaking, optimal solution to an SVM classifier is associated with finding the maximum margin. For a soft-margin SVM, support vectors can lie on the margins as well as inside the margins. Those support vectors that lie in between the separating boundaries and the margins represent misclassified samples from the training set.

Let prelabeled training sample pairs be $(\mathbf{x}_1, d_1), \ldots, (\mathbf{x}_N, d_N)$, where $\mathbf{x}_i$ are the input vectors and d_i are the corresponding labels in a binary class $Y = \{-1, +1\}$. A soft-margin SVM is described by a set of equations:

$$d_i(\langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \ i = 1, 2, \ldots, N, \tag{14}$$

or as a quadratic optimization problem:

$$\min_{\mathbf{w}, b, \xi} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right), \tag{15}$$

where $C \geq 0$ is a user-specified *regularization parameter* that determines the trade-off between the upper bound on the complexity term and the empirical error (or, training error), and ξ_i are the slack variables. C can be determined empirically, for example, though a cross-validation process. This leads to the primal form of the Lagrangian optimization problem for a soft-margin SVM,

$$L(\mathbf{w}, b, \alpha, \xi, \mu) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i$$
$$- \sum_{i=1}^{N} \alpha_i [d_i(\langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \rangle + b) - 1 + \xi_i], \tag{16}$$

where μ_i is the Lagrange multiplier to enforce the positivity of ξ_i. Solutions to the soft-margin SVM can be obtained by solving the dual problem of (16),

$$\max_{\alpha} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j d_i d_j k(\mathbf{x}_i, \mathbf{x}_j) \right\}$$

subject to $\quad \sum_{i=1}^{N} \alpha_i d_i = 0, \quad C \geq \alpha_i \geq 0, \ i = 1, 2, \ldots, N. \tag{17}$

The KKT (Karush-Kuhn-Tucker) conditions for the primal optimization problem [31] are

$$\alpha_i [d_i(\langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_i) \rangle + b) - 1 + \xi_i] = 0, \quad \mu_i \xi_i = 0, \quad i = 1, 2, \ldots, N. \tag{18}$$

The KKT conditions are associated with the necessary and in some cases sufficient conditions for a set of variables to be optimal. The Lagrangian multipliers α_i are nonzero only when the KKT conditions are met.

The corresponding vectors $\mathbf{x}_i$, for nonzero α_i, satisfying the 1st set of equations in (18) are the so-called *support vectors*. Once α_i are determined, the optimal solution $\mathbf{w}$ can be obtained as

$$\mathbf{w} = \sum_{i=1}^{N_s} \alpha_i d_i \mathbf{\Phi}(\mathbf{x}_i), \tag{19}$$

where $\mathbf{x}_i \in \mathrm{SV}$, and the bias b can be determined by the KKT conditions in (18).

Instead of solving the primal problem in (16), one usually solve the equivalent dual problem of a soft-margin SVM described as

$$\max_{\alpha} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j d_i d_j k(\mathbf{x}_i, \mathbf{x}_j) \right\} \tag{20}$$

subject to $\quad \sum_{i=1}^{N} \alpha_i d_i = 0, \quad C \geq \alpha_i \geq 0, \ i = 1, 2, \ldots, N$

noting that the slack variables ξ_i and the weight vector $\mathbf{w}$ do not appear in the dual form. Finally, the decision function
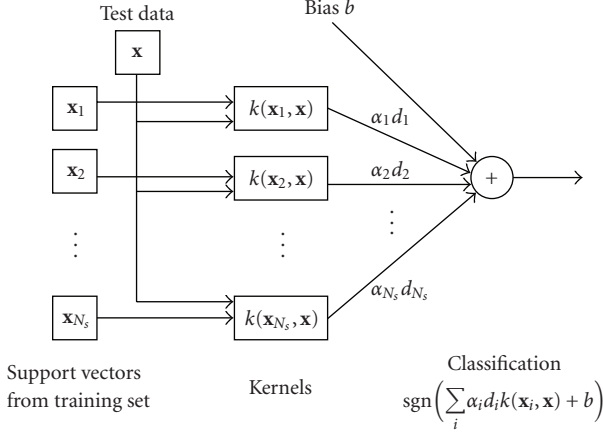
FIGURE 16: Block diagram of an SVM classifier for a two-class case.

for a soft-margin SVM classifier is

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{N_s} \alpha_i d_i k(\mathbf{x}, \mathbf{x}_i) + b\right), \qquad (21)$$

where N_s is the number of support vectors $\mathbf{x}_i$, $\mathbf{x}_i \in \text{SV}$, and $k(\mathbf{x}, \mathbf{x}_i) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_i) \rangle$ is the selected kernel function. Solution to (20) can be obtained as a (convex) quadratic programming (QP) problem [32]. Figure 16 shows the block diagram of an SVM classifier where $\mathbf{x}$ is the input vector drawn from the test set, and $\mathbf{x}_i$, $i = 1, \ldots, N_s$, are the support vectors drawn from the training set. It is worth noting that although the block diagram structure of the SVM looks similar to an RBF neural network, the fundamental differences exist. First, the weights $\alpha_i d_i$ to the output of an SVM contain Lagrange multipliers α_i that are associated with the solutions of the constrained optimization problem in the SVM learning, such that the generalization error (or the classification error on the test set) is minimized. In RBF neural networks, the weights are selected for minimizing the empirical error (i.e., the error between the desired outputs d and the network outputs from the training set). Second, for a general multilayer neural network, activation functions are used in the hidden layer to achieve the nonlinearity, while an SVM employs a kernel function $k(\cdot)$ only dependent on the difference of $\mathbf{x}$ from the test set and the support vectors from the training set.

5.4. Radial basis function kernel C-SVMs for classification of power disturbances

The proposed SVMs use Gaussian RBF (radial-basis function) kernels, that is,

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{\gamma}\right). \qquad (22)$$

For a C-SVM with RBF kernels, there are two parameters to be selected: one is the regularization parameter C in (16), an-

TABLE 2: AND problem for classifying voltage dips due to faults.

Input vector $\mathbf{x}$	Desired response d
(1,1)	+1
(1,0)	−1
(0,0)	−1
(0,1)	−1

other is the kernel parameter γ. An RBF kernel is chosen since it can behave like a linear kernel or a sigmoid kernel under different parameter settings. An RBF is also known to have less numerical difficulties since the kernel matrix $\mathbf{K} = [k_{ij}]$, $i, j = 1, \ldots, N$, is symmetric positive definite. The parameters (C, γ) should be determined prior to the training process. Cross validation is used for the purpose of finding the best parameters (C, γ). We use the simple grid-search method described in [33]. First, a coarse grid-search is applied, for example, $C = \{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$, $\gamma = \{2^{15}, 2^{13}, \ldots, 2^{-3}\}$. The search is then tuned to a finer grid in the region where the predicted error rate from the cross validation is the lowest in the coarse search. Once (C, γ) are determined, the whole training set is used for the training. The learned SVM with fixed weights is then ready to be used as a classifier for samples $\mathbf{x}$ drawn from the test set. Figure 17 describes the block diagram of the entire process.

It is worth mentioning that normalization is applied before each feature vector before it is fed into the classifier either for training or classification. Normalization is applied to each component (or subgroup of components) in a feature vector, so that all components have the same mean and variance. Noting that the same normalization process should be applied to the training and test sets.

5.5. Example 1

To show how an SVM can be used for classification of power system disturbances, a very simple hypothetical example is described of classifying fault-induced voltage dips.

A voltage dip is defined as an event during which the voltage drops below a certain level (e.g., 90% of nominal). Furthermore, for dips caused by faults, RMS voltage versus time is close to a rectangular shape. However, a sharp drop followed by a slow recovery is caused by transformer saturation or induction motor starting depending on whether they have a similar voltage dropping in all three phases.

In the ideal case, classification of fault-induced voltage dips can be described as an AND problem. Let each feature vector $\mathbf{x} = [x_1, x_2]^T$ consist of two components, where $x_1 \in \{0, 1\}$ indicates whether there is a fast drop in RMS voltage to below 0.9 pu (i.e., rectangular shaped drop). If there is a sharp voltage drop to below 0.9 pu, then x_1 is set to 1. $x_2 \in \{0, 1\}$ indicates whether there is a quick recovery in RMS voltage (i.e., rectangular shaped recovery in the RMS sequence). When the recovery is fast, then x_2 is set to 1.

Also, define the desired output for dips due to faults as $d = 1$, and $d = -1$ for all other dips. Table 2 describes such an AND operation.
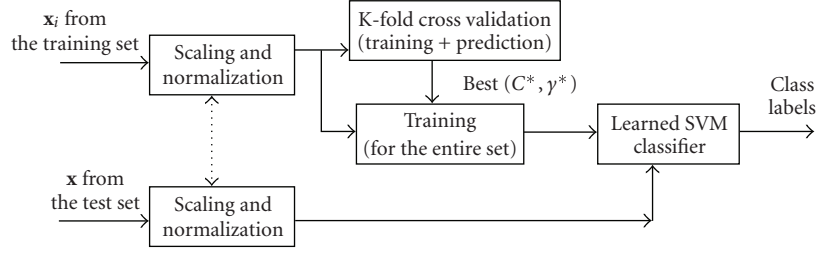
FIGURE 17: Block diagram for learning an RBF kernel SVM. The SVM is subsequently used as a classifier for input **x** from a test set.

However, a voltage drop or recovery will never be absolutely instantaneous, but instead always take a finite time. The closeness to rectangularity is a fuzzy quantity that can be defined as a continuous function ranging between 0 and 1.0 (e.g., by defining a fuzzy function, see [34]), where 0 implies a complete flat shape while 1.0 an ideal rectangular shape ($\theta = 90^\circ$). Consequently, each component of the feature vector **x** takes a real value within $[0, 1]$.

In this simple example, the training set contains the following 12 vectors:

$$\mathbf{x}_1 = (1.0, 1.0), \quad \mathbf{x}_2 = (1.0, 0.0), \quad \mathbf{x}_3 = (0.0, 0.0),$$
$$\mathbf{x}_4 = (0.0, 1.0), \quad \mathbf{x}_5 = (0.8, 0.7), \quad \mathbf{x}_6 = (0.9, 0.4),$$
$$\mathbf{x}_7 = (0.2, 0.8), \quad \mathbf{x}_8 = (0.4, 0.8), \quad \mathbf{x}_9 = (0.8, 0.9),$$
$$\mathbf{x}_{10} = (0.9, 0.5), \quad \mathbf{x}_{11} = (0.3, 0.2), \quad \mathbf{x}_{12} = (0.3, 0),$$
$$\tag{23}$$

and their corresponding desired outputs are

$$d_1 = d_5 = d_9 = +1,$$
$$d_2 = d_3 = d_4 = d_6 = d_7 = d_8 = d_{10} = d_{11} = d_{12} = -1.$$
$$\tag{24}$$

Figure 18 shows the decision boundaries obtained from training an SVM using Gaussian RBF kernels with two different γ values (related to the spread of kernels). One can see that the bending of the decision boundary (i.e., the solid line) changes depending on the choice of the parameter related to Gaussian spread. In Figure 18(a), the RBF kernels have a relatively large spread and the decision boundary becomes very smooth which is close to a linear boundary. While in Figure 18(b) the RBFs have a smaller spread and the decision boundary has a relatively large bending (nonlinear boundary). If the features of event classes are linearly nonseparable, a smoother boundary would introduce more training errors. Interestingly enough, the decision boundary obtained from the training set (containing feature vectors with soft values of rectangularity in voltage drop and recovery) divides the feature space into two regions where the up-right region (notice that the shape of region makes good sense!) is associated with the fault-induced dip events.

When the training set contains more samples, one may expect an improved classifier. Using more training data will lead to changes in the decision boundary, the margins, and the distance between the margins and the boundary. We

might also find some support vectors that are lying between the boundary and the margin, indicating training errors. However, for achieving a good performance on the test set, some training errors have to be allowed.

It is worth mentioning that, to apply this example in real applications, one should first compute the RMS sequence from a given power quality disturbance recording, followed by extracting these features from the shape of RMS sequence. Further, using features instead of directly using signal waveform for classification is a common practice in pattern classification systems. Advantages of using features instead of signal waveform include, for example, reducing the complexity of classification systems, exploiting the characteristics that are not obvious in the waveform (e.g., frequency components), and using disturbance sequences without time warping.

5.6. Example 2

A more realistic classifier has been developed to distinguish between five different types of dips. The classifier has been trained and tested using features extracted from the measurement data obtained from two different networks.

The following five classes of disturbances are distinguished (see Table 3):

(D1) voltage dips with a main drop in one phase;
(D2) voltage dips with a main, identical, drop in two phases;
(D3) voltage dips due to three phase faults;
(D4) voltage dips with a main, but different, drop in two phases;
(D5) voltage disturbances due to transformer energizing.

Each measurement recording includes a short prefault waveform (approximately 2 cycles long) followed by the waveform containing a disturbance. The nominal frequency of the power networks is 50 Hz and the sampling frequencies are 1000 Hz, 2000 Hz, or 4800 Hz. The data were originated from two different networks from two European countries A and B.

Synthetic data have been generated by using the power network simulation toolbox "SimPowerSystems" in Matlab. The model is a 11-kV network consisting of a voltage source, an incoming line feeding four outgoing lines via a busbar. The outgoing lines are connected to loads with different load characteristics in terms of active and reactive power. At one of these lines, faults are generated. In order to simulate different
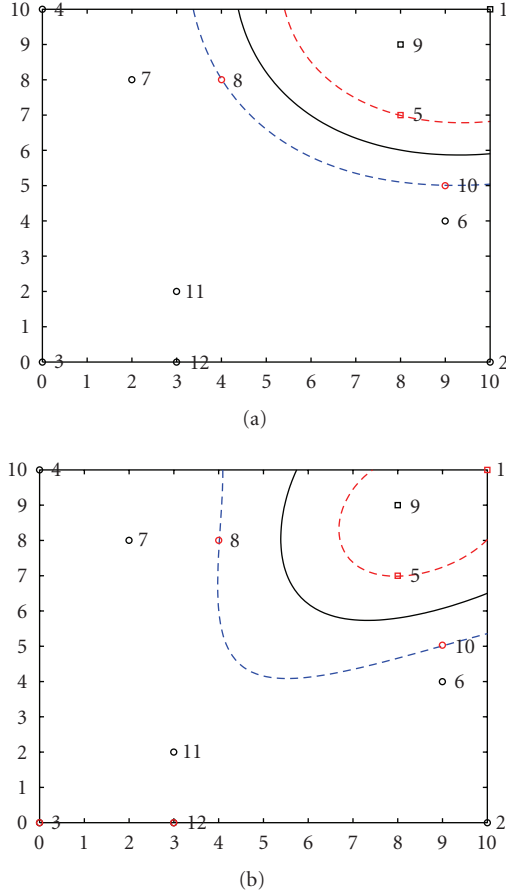
(a)

(b)

FIGURE 18: The decision boundaries and support vectors obtained by training an SVM with Gaussian RBF kernel functions. Solid line: the decision boundary; dashed lines: the margins. (a) The parameter for RBF kernels is $\gamma = 36m_0$; (b) the parameter for RBF kernels is $\gamma = 8m_0$. For both cases, the regularization parameter is $C = 10$, and $m_0 = 2$ is the size of the input vector. Solid line is the decision boundary. Support vectors are the training vectors that are located on the margins (the dotted lines) and in between the boundary and the margin (misclassified samples). For the given training set, there is no misclassified samples. The two sets of feature vectors are marked with rectangular and circles, respectively. Notice that only support vectors are contributed to the location of the decision boundary and margins. In the figure, all feature values are scaled by a factor of 10.

TABLE 3: Number of voltage recordings used.

Disturbance type	Network A	Network B	From synthetic
D1	141	471	225
D2	181	125	225
D3	251	14	223
D4	127	196	250
D5	214	0	0

waveforms, the lengths of the incoming line and the outgoing lines as well as the duration of the disturbances are randomly varied within given limits. Finally, the voltage disturbances are measured at the busbar.

TABLE 4: Classification results using training and test data from network A.

	D1	D2	D3	D4	D5	Nonclassified	Detection rate
D1	63	0	1	3	0	4	88.7%
D2	0	84	4	0	0	3	92.3%
D3	5	3	113	0	0	5	89.7%
D4	0	0	0	63	1	0	98.4%
D5	0	0	0	0	103	4	96.3%

TABLE 5: Classification results for data from network B, where the SVM is trained with data from network A.

	D1	D2	D3	D4	Nonclassified	Detection rate
D1	463	1	0	1	6	99.1%
D2	0	118	0	0	7	94.4%
D3	2	1	11	0	0	78.6%
D4	0	1	0	187	8	95.4%

Segmentation is first applied to each measurement recording. Data in the event segment after the first transition segment are used for extracting features. The features are based on RMS shape and the harmonic energy. For each phase, the RMS voltage versus time is computed. Twenty feature components are extracted from equally distance-sampling the RMS sequence, starting from the triggering point of the disturbance. Further, to include possible differences of harmonic distributions in the selected classes of disturbances, some harmonic-related features are also used. Four harmonic-related feature components are extracted from the magnitude spectrum from the data in the event segment. They are the magnitudes of 2nd, 5th, and 9th harmonics and the total harmonics distortion (THD) with respect to the magnitude of power system fundamental frequency component. Therefore, for each three-phase recording, the feature vector consists of 72 components. Feature normalization is then applied to ensure that all components are equally important. It is worth mentioning that since there are 60 feature components related to the 3 RMS sequences and 12 feature components related to harmonic distortions, it implies that more weight has been put on the shape of RMS voltage as compared to harmonic distortion.

C-SVM with RBF kernels is chosen. Grid search and a 3-fold validation are used for determining the optimal parameters (C, γ) for the SVM. Further, we use separate SVMs for 5 different classes, each classifying between one particular type and the remaining. These SVMs are connected in a binary tree structure.

Experiment results

Case 1. For the first case, data from the network A are split into two subsets, one is used for training, and the other is

Table 6: Classification results for data from network A, where the SVM is trained with data from network B.

	D1	D2	D3	D4	Nonclassified	Detection rate
D1	137	0	2	1	1	97.2%
D2	0	154	16	0	11	85.1%
D3	10	1	232	0	8	92.4%
D4	4	1	1	121	0	95.2%

Table 7: Classification results for data from network B, where the SVM is trained with data from network A and synthetic data.

	D1	D2	D3	D4	Nonclassified	Detection rate
D1	473	1	0	0	1	99.5%
D2	0	122	2	0	1	97.6%
D3	0	0	187	0	9	95.4%
D4	15	31	0	14	7	20.0%

used for testing. The resulting classification results are shown in Table 4. The resulting overall detection rate is 92.8% showing the ability of the classifier to distinguish between the classes by using the chosen features.

Case 2. As a next case, all data from network A have been used to train the SVM, which was next applied to classify the recordings obtained for network B. The classification results are presented in Table 5. The resulting overall detection rate is 96.1%. The detection rate is even higher than in Case 1 were training and testing data were obtained from the same network. The higher detection rate may be due to the larger training set being available. It should also be noted that the detection rate is an estimation based on a small number of samples (the misclassifications), resulting in a large uncertainty. Comparison of the detection rates for different classifiers should thus be treated with care. But despite this, the conclusion can be drawn that the SVM trained from data in one network can be used for classification of recordings obtained from another network.

Case 3. To further test the universality of a trained SVM, the rules of network A and network B were exchanged. The classifier was trained from the data obtained in network B and next applied to classify the recordings from network A. The results are shown in Table 6. The resulting overall detection rate is still high at 92.0% but somewhat lower than in the previous case. The lower detection rate is mainly due to the larger number of nonclassifications and incorrect classifications for disturbance types D2 and D3.

Case 4. As a next experiment, synthetic data were used for training the SVM. Using 100% synthetic data for training resulted in a low detection rate, but mixing some measurement data with the synthetic data gave acceptable results. Table 7 gives the performance of the classifier when the training set consists for 25% of data from network A and for the remain-

ing 75% of synthetic data. The resulting overall detection rate is 92.1%, note; however the very bad performance for class D4.

6. CONCLUSIONS

A significant amount of work has been done towards the development of methods for automatic classification of power-quality disturbances. The current emphasis in literature is on statistical methods, especially artificial neural networks. With few exceptions, most existing work defines classes based on disturbance types (e.g., dip, interruption, and transient), rather than classes based on their underlying causes. Such work is important for the development of methods but has, as yet, limited practical value. Of more practical needs are tools for classification based on underlying causes (e.g., dip due to motor start, transient due to synchronized capacitor energizing). A further limitation of many of the studies is that both training and testing are based on synthetic data. The advantages of using synthetic data are understandable, both in controlling the simulations and in overcoming the difficulties of obtaining measurement data. However, the use of synthetic data further reduces the applicability of the resulting classifier.

A number of general issues should be considered when designing a classifier, including segmentation and the choice of appropriate features. Especially the latter requires insight in the causes of power-system disturbances and the resulting voltage and current waveforms. Two segmentation methods are discussed in this paper: the Kalman-based method applied to voltage waveforms; and a derivative-based method applied to RMS sequences.

Two classifiers are discussed in detail in this paper. A rule-based expert system is discussed that allows classification of voltage disturbances into nine classes based on their underlying causes. Further subclasses are defined for some of the classes and additional information is obtained from others. The expert system makes heavy use of power-system knowledge, but combines this with knowledge on signal segmentation from speech processing. The expert system has been applied to a large number of measured disturbances with good classification results. An expert system can be developed without or with a limited amount of data. It also makes more optimal use of power-system knowledge than the existing statistical methods.

A support vector machine (SVM) is discussed as an example of a robust statistical classification method. Five classes of disturbances are distinguished. The SVM is tested by using measurements from two power networks and synthetic data. An interesting and encouraging result from the study is that a classifier trained by data from one power network gives good classification results for data from another power network. It is also shown that training using synthetic data does not give acceptable results for measured data, probably due to using less realistic models in generating synthetic data as compared with real data. Mixing measurements and synthetic data improves the performance but some poor performance remains.

A number of challenges remain on the road to a practically applicable automatic classifier of power-quality disturbances. Feature extraction and segmentation still require further attention. The main question to be addressed here is: what information remains hidden in power-quality disturbance waveforms and how can this information be extracted?

The advantage of statistical classifiers is obvious as they optimize the performance, which is especially important for linearly nonseparable classes, as is often the case in the real world. However, in the existing volume of work on statistical classifiers it has shown to be significantly easier to include power-system knowledge in an expert system than in a statistical classifier. This is partly due to the nonoverlapping knowledge between the two research areas, but also due to the lack of suitable features. More research efforts should be aimed at incorporating power-system knowledge in statistical classifiers.

The existing classifiers are mainly limited to the classification of dips, swells, and interruptions (so-called "voltage magnitude events" or "RMS variations") based on voltage waveforms. The work should be extended towards classification of transients and harmonic distortion based on underlying causes. The choice of appropriate features may be an even greater challenge than for voltage magnitude events. Also, more use should be made of features extracted from the current waveforms.

An important observation made by the authors was the strong need for large amounts of measurement data obtained in power networks. Next to that, theoretical and practical power-system knowledge and signal processing knowledge are needed. This calls for a close cooperation between power-system researchers, signal processing researchers, and power network operators. All these issues make the automatic classification of power-quality disturbances a sustained interesting and challenging research area.

REFERENCES

[1] A. K. Khan, "Monitoring power for the future," *Power Engineering Journal*, vol. 15, no. 2, pp. 81–85, 2001.

[2] M. McGranaghan, "Trends in power quality monitoring," *IEEE Power Engineering Review*, vol. 21, no. 10, pp. 3–9, 21, 2001.

[3] M. H. J. Bollen, *Understanding Power Quality Problems: Voltage Sags and Interruptions*, IEEE Press, New York, NY, USA, 1999.

[4] L. Angrisani, P. Daponte, and M. D'Apuzzo, "Wavelet network-based detection and classification of transients," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 5, pp. 1425–1435, 2001.

[5] M. H. J. Bollen and I. Y. H. Gu, *Signal Processing of Power Quality Disturbances*, IEEE Press, New York, NY, USA, 2006.

[6] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, "Power disturbance classifier using a rule-based method and wavelet packet-based hidden Markov model," *IEEE Transactions on Power Delivery*, vol. 17, no. 1, pp. 233–241, 2002.

[7] P. K. Dash, S. Mishra, M. M. A. Salama, and A. C. Liew, "Classification of power system disturbances using a fuzzy expert system and a Fourier Linear Combiner," *IEEE Transactions on Power Delivery*, vol. 15, no. 2, pp. 472–477, 2000.

[8] Z.-L. Gaing, "Implementation of power disturbance classifier using wavelet-based neural networks," in *IEEE Bologna PowerTech Conference*, vol. 3, p. 7, Bologna, Italy, June 2003.

[9] Z.-L. Gaing, "Wavelet-based neural network for power disturbance recognition and classification," *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1560–1568, 2004.

[10] A. M. Gaouda, M. M. A. Salama, M. R. Sultan, and A. Y. Chikhani, "Power quality detection and classification using wavelet-multiresolution signal decomposition," *IEEE Transactions on Power Delivery*, vol. 14, no. 4, pp. 1469–1476, 1999.

[11] J. Huang, M. Negnevitsky, and D. T. Nguyen, "A neural-fuzzy classifier for recognition of power quality disturbances," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 609–616, 2002.

[12] S.-J. Huang, T.-M. Yang, and J.-T. Huang, "FPGA realization of wavelet transform for detection of electric power system disturbances," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 388–394, 2002.

[13] M. Kezunovic and Y. Liao, "A novel software implementation concept for power quality study," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 544–549, 2002.

[14] C. H. Lee and S. W. Nam, "Efficient feature vector extraction for automatic classification of power quality disturbances," *Electronics Letters*, vol. 34, no. 11, pp. 1059–1061, 1998.

[15] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier—part 1: theoretical foundation," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 222–228, 2000.

[16] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier—part 2: application," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 229–235, 2000.

[17] S. Santoso, J. D. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt, "A scalable PQ event identification system," *IEEE Transactions on Power Delivery*, vol. 15, no. 2, pp. 738–743, 2000.

[18] S. Santoso and J. D. Lamoree, "Power quality data analysis: from raw data to knowledge using knowledge discovery approach," in *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, vol. 1, pp. 172–177, Seattle, Wash, USA, July 2000.

[19] E. Styvaktakis, M. H. J. Bollen, and I. Y. H. Gu, "Expert system for classification and analysis of power system events," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 423–428, 2002.

[20] E. Styvaktakis, *Automating power quality analysis*, Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 2002.

[21] M. Wang and A. V. Mamishev, "Classification of power quality events using optimal time-frequency representations—part 1: theory," *IEEE Transactions on Power Delivery*, vol. 19, no. 3, pp. 1488–1495, 2004.

[22] M. Wang, G. I. Rowe, and A. V. Mamishev, "Classification of power quality events using optimal time-frequency representations—part 2: application," *IEEE Transactions on Power Delivery*, vol. 19, no. 3, pp. 1496–1503, 2004.

[23] J. V. Wijayakulasooriya, G. A. Putrus, and P. D. Minns, "Electric power quality disturbance classification using self-adapting artificial neural networks," *IEE Proceedings: Generation, Transmission and Distribution*, vol. 149, no. 1, pp. 98–101, 2002.

[24] A. M. Youssef, T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Disturbance classification utilizing dynamic time warping classifier," *IEEE Transactions on Power Delivery*, vol. 19, no. 1, pp. 272–278, 2004.

[25] J. Goldberger, D. Burshtein, and H. Franco, "Segmental modeling using a continuous mixture of nonparametric models," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 262–271, 1999.

[26] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[27] R. G. Cowell, S. L. Lauritzen, and D. J. Spiegelhater, *Probabilistic Networks and Expert Systems*, Springer, New York, NY, USA, 2nd edition, 2003.

[28] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, UK, 2004.

[29] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[30] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.

[31] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[32] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, 1995.

[33] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," *LIBSVM—A library for Support Vector Machines*, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[34] B. Kosko, *Fuzzy Engineering*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.

Math H. J. Bollen is the Manager of power quality and EMC at STRI AB, Ludvika, Sweden, and a Guest Professor at EMC-on-Site, Luleå University of Technology, Skellefteå, Sweden. He received the M.S. and Ph.D. degrees from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1985 and 1989, respectively. Before joining STRI in 2003, he was a Research Associate at Eindhoven University of Technology, a Lecturer at University of Manchester Institute of Science and Technology, Manchester, UK, and Professor in electric power systems at Chalmers University of Technology, Gothenburg, Sweden. His research interests cover various aspects of EMC, power quality and reliability, and related areas. He has published a number of fundamental papers on voltage dip analysis and has authored two textbooks on power quality.

Irene Y. H. Gu is a Professor in signal processing at the Department of Signals and Systems at Chalmers University of Technology, Sweden. She received the Ph.D. degree in electrical engineering from Eindhoven University of Technology, The Netherlands, in 1992. She was a Research Fellow at Philips Research Institute IPO, The Netherlands, and Staffordshire University, UK, and a Lecturer at The University of Birmingham, UK, during 1992–1996. Since 1996, she has been with Chalmers University of Technology, Sweden. Her current research interests include signal processing methods with applications to power disturbance data analysis, signal and image processing, pattern classification, and machine learning. She served as an Associate Editor for the IEEE Transactions. on Systems, Man, and Cybernetics during 2000–2005, the Chair-Elect of Signal Processing Chapter in IEEE Swedish Section 2002–2004, and is a Member of the editorial board for EURASIP Journal on advances in Signal Processing since July 2005. She is the Coauthor of the book *Signal Processing of Power Quality Disturbances* published by Wiley/IEEE Press 2006.

Peter G. V. Axelberg received the M.S. and Tech. Licentiate degrees from Chalmers University of Technology, Gothenburg, Sweden, in 1984 and 2003, respectively. From 1984 to 1992, he was at ABB Kabeldon in Alingsås, Sweden. In 1992, he cofounded Unipower where he is currently active as the Manager of business relations and research. Since 1992, he has also been a Lecturer at University College of Borås, Sweden. His research activities are focused on power quality measurement techniques.

Emmanouil Styvaktakis received his B.S. degree in electrical engineering from the National Technical University of Athens, Greece, in 1995, M.S. degree in electrical power engineering from the Institute of Science and Technology, University of Manchester, in 1996, and Ph.D. degree in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 2002. He is currently with the Hellenic Transmission System Operator (HTSO). His research interests are power quality and signal processing applications in power systems.