A unified approach for determining the underlying causes of nonstationary disturbances

Praveen Pankajakshan*
INRIA Sophia Antipolis - Méditerranée,
2004 route des lucioles, B.P.93,
06902 Sophia-Antipolis Cedex, France
E-mail: praveenpankaj@ieee.org
*Corresponding author

Abstract: In this paper we have developed a framework for automatic detection and classification of disturbances in a distribution feeder. Unlike existing techniques, we examine these disturbances based on their underlying causes. A fundamental and important part of this process is the segmentation of the disturbances from the captured signal using either a Kalman Filter (KF) or a Multi-resolution Signal Decomposition (MSD) technique. The segmentation process divides the quasi-stationary Root-Mean-Square (RMS) signal into pre-disturbance, disturbance and post-disturbance regions. The pre- and post-disturbance segments are essentially stationary while the nonstationary nature is extracted as the disturbance segment. An important facet that is often forgotten is the morphology of the captured voltage or current signal's RMS. We represent each of the segmented regions as *strings* or *sequences* of predefined wave patterns, called *primitives*. Any syntactically correct combination of these *primitives* will define the morphology of the *mother RMS signal*. The grammar and the model for each class is built from a set of positive examples (I_+) or the *learning set* using the Error-correcting Grammatical Inference (ECGI) learning technique. A stochastic extension of the ECGI similar to a Viterbi-like dynamic algorithm in combination with the *k-nearest neighbor algorithm (kNN)* classifier is used to recognize the captured data. The structure of a new pattern can be learnt by updating the current grammar with the inferred production rules.

Keywords: power disturbance, power quality, Tap Changing Under Load (TCUL), wavelets, Kalman filter, Error Correcting Grammatical Inference (ECGI), k-nearest neighbor (kNN).

Reference to this paper should be made as follows: P. Pankajakshan (2008) 'A unified approach for determining the underlying causes of nonstationary disturbances', Int. J. Computer Applications in Technology , Vol. 1, Nos. 1, pp.xx–yy.

Biographical notes: Praveen Pankajakshan received his B.Tech. degree in the Electrical Engineering from the India Institute of Technology Roorkee and his Masters degree in Electrical and Computer Engineering from Texas A&M University, College Station, in 2005. He is currently pursuing his Ph.D. in Computer Science at INRIA Sophia-Antipolis, France. His research interests are primarily in the area of statistical signal and image processing, inverse problems, and pattern recognition.

1 INTRODUCTION

The burgeoning need to reduce maintenance costs and to provide for ways to improve the stability and quality has led the utilities to focus on long-term recording and analysis of power system disturbances. These disturbances include events that are normal-low priority, abnormal -high priority, and disturbances that have never been observed (low or high priority). Any power line disturbance that affects the performance of sensitive electronic equipment is related to *Power Quality* (Gaouda et al. (2000);Heydt

(1998)). Recording of a power system disturbance includes both time and frequency domain conducted parameters, and the disturbances may be caused due to over/under voltages, interruptions, Capacitor, transformer energizing, Tap Changing Under Load (TCUL), Motor/load switching, power electronic devices operating, transients, phase imbalance, frequency aberrations, re-strikes, blown fuses, arcing, storm and environment related damages or others.

However, analysts frequently face the situation where they are flooded with a large amount of data. Even a momentary interference of two seconds requires about two hours of manual data sorting. This is not only time consuming but often repetitive. The research towards the automatic processing of the data is inspired by the need to overcome this burden. Existing technologies examine the captured data as *sags*, *swells* or *harmonics*, and there has never been an effort to individually identify the underlying cause of these disturbances. Identification of the cause and defining the priority handling levels are very important for an engineer or technician to isolate the problem early, and to find an immediate solution. In this paper, we solve the issue of automatic classification of disturbances in a distribution system and also facilitate the conversion of the large amount of data to information that will allow immediate assessment.

A number of ideas have been introduced for simplifying the task of users who study patterns in waveforms. The *TimeSearcher* (Hochheiser and Shneiderman (2002)) is one such approach that finds patterns in complex time-series data based on a ROI query. Another important application, in the lines of this paper has been discussed for automotive fault diagnostics (Crossman et al. (2003)) and in parsing EKG signal (Morrill (1998)).

This paper is organized in the following manner. In Section 2, we describe briefly the experimental set up and the disturbances that are observable on a typical distribution feeder circuit. We employ two different approaches for the localization and detection of disturbances in the voltage and current signal. In Section 3.2, we segment the signal either with a recursive optimal estimator like the Kalman Filter (KF) or using the Multi-resolution Signal Decomposition (MSD) technique. Formulating the KF requires knowledge of the initial conditions, the process noise, and the measurement noise. The estimated state variables, on completion of the initialization cycle are used as the initial state variables for the successive cycle. When all the state variables reach steady state, under nonstationary conditions, the measurement residual contains outliers that is used for segmenting the nonstationary region. The MSD on the other hand decomposes the original signal into a smoothed and detailed version of the signal. We can localize the disturbance by enhancing the squared detail Daubechies' wavelets (Daub4) (Mallat (1999)). A peak detection algorithm detects the outliers and hence the nonstationary regions in the signal. We provide valuable arguments to support the use of MSD over the KF for signal segmentation.

In Section 4 we use the compactness, bound and clustering ability as guidelines to define the desired features. The generic features are inferred from time domain or statistical analysis of the signal. Other features are derived from either the segmentation or the morphology of the signal.

A syntactic approach to recognizing disturbances involves partitioning the Root-Mean-Square (RMS) of the signal into *sub-patterns* or *primitives* and defining a structural relationship among them. We tackle the difficult problem of assigning each piecewise linear signal segment to the right primitives in Section 5.2. Primitive templates are first extracted from the training data and these are then stored as reference signatures. In Section 5.2.4, we generate the set of rules, the *grammar* for connecting the *primitives*. The production rules are derived from a finite set of positive training strings I_+. If the new string is not acceptable by the grammar, a standard error-correcting parsing (Rulot and Vidal (1988);Rulot et al. (1989);Vidal et al. (1993)), based on similarity of patterns is used to determine the distance measure between a sentence representing a known pattern and a sentence representing an unknown pattern. Then, every string that cannot be parsed is adopted to determine a string in the language of the current grammar that is error-correcting closest to the input string. Since each step has more than one production rules for deriving it, probabilities are assigned to alternatives in order to indicate likelihood of occurrence or to guide the process of selection. This stochastic recognition algorithm is used in combination with the *k-Nearest Neighbor (kNN)* features by the classifier for recognizing the disturbances and their cause.

Once the normal disturbances are classified, the unknown or abnormal disturbances have to be analyzed to detect emerging fault conditions. Classification becomes easier if the disturbance exhibits characteristics or features similar to a known abnormal disturbance. In Section 7, we provide a case study that shows the complexity and difficulty in concluding some disturbances as abnormal. Temporal aggregation and correlation over time is proposed for detecting emerging or existing conditions.

2 DATA COLLECTION, MANAGEMENT AND PROCESSING

2.1 Experimental Setup and Management

Disturbance monitors are instruments designed to detect and record data on power system variations. But only the objective for a particular project will determine the choice of monitoring equipment, the method of collecting data, the triggering thresholds needed, the data analysis technique to employ, and the overall level of effort required of the project.

The monitoring units should store the high-speed waveform data only when any one of the power system parameter crosses the configured thresholds (IEEE (1995)) that are set for it. The thresholds are set below (more sensitive) equipment susceptibility levels to ensure that all the disturbances are recorded. There is always a trade-off be-

tween being overly sensitive and capturing the slight fluctuations, and not capturing the *important* disturbances. A server coordinates the activities, manages the long-term data storage, and communicates with the external clients. The preliminary calculations are handled by the server, while a more detailed and advanced analysis or classification is carried out at the client level. In the remaining part of this paper, our discussion will be on the analysis at the client level that involves the integration of the recognition algorithm.

2.2 Description of the recorded data

The power signal contains the fundamental frequency (f) and the higher order harmonics that are an integral multiple of f. To prevent aliasing, and to satisfy the *Nyquist criterion*, the sampling frequency (f_s) is:

$$f_s \geq 2f_h \tag{1}$$

where, f_h is the highest frequency in the signal. For our example, f_s is assumed to be 15360 samples per second (i.e., 256 samples per cycle at $f = 60$ Hz). The signal that is used by all the algorithms are either the voltage (three phases) or current waveform (three phases and a neutral) or the RMS of the signals (see Section 3.1.2).

2.3 Disturbances in a Distribution System

The disturbances that occur in a distribution or a transmission system at a higher level could be classified into one of the following:

- normal Disturbance, (every day observable, low priority),
- abnormal Disturbance, (requires immediate attention, high priority), and
- unknown, (low or high priority).

The principal phenomena that causes these electromagnetic disturbances, as classified by the IEC is listed in (IEEE (1995)). The normal disturbances in a Power System constitute the majority of the data that occurs in a Power System. It is very important that the automatic recognition system is able to identify these disturbances and classify them right (no false positives), so that it is easy to isolate those which might affect the reliability of the system. These include:

- Capacitor Energizing,
- Capacitor De-energizing,
- TCUL Up,
- TCUL Down,
- Load Variations (Increase and Decrease),
- Inrush (e.g. caused by lightning), and
- Overcurrent Faults (Normal).

The abnormal disturbances represent devices that may have failed or is in the process of failing in which case they would provide early warning signals about the impending failure. Examples, are failing TCUL or equipment, a stuck capacitor switch, etc. It is very likely that the algorithm cannot enumerate and characterize all possible operating modes of the power system devices. In such a case, the algorithm categorizes these disturbances as *unknown*.

Each of the above disturbances exhibit a particular characteristic that is useful for recognizing their sources. For example, a capacitor switching disturbance on a distribution system would exhibit a step change on the RMS of the current or the voltage accompanied by switching transients (Bollen et al. (2007)). Similarly, a voltage sage is very commonly associated with a Motor coming into operation. So the identification of the shape of the RMS of the waveform is very important. Some of the commonly observed RMS signal shapes are:

1. Step up in current (voltage),
2. Step down in current (voltage),
3. Short duration current (voltage) sag,
4. Short duration current (voltage) swell,
5. Fault induced current swell,
6. Fault induced voltage sag,
7. Impulsive current (voltage) transient, and
8. Short or long plateau of the current (voltage).

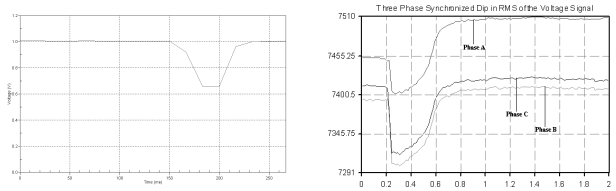
2.3.1 Transients

Transient pertains to or designates a phenomenon that varies between two consecutive steady states during a time interval that is short compared to the time scale of interest (see Fig. 3 and Fig. 4(b)). A transient can be a unidirectional *impulse* of either polarity or a damped *oscillatory* wave with the first peak occurring in either polarity. *Impulsive transients* causes a sudden non-power frequency change in the steady-state condition of voltage or current that is unidirectional in polarity (primarily either positive or negative). *Oscillatory transient* are a sudden, non-power frequency change in the steady-state condition of voltage or current that includes both positive and negative polarity value.

2.3.2 Short duration power disturbances

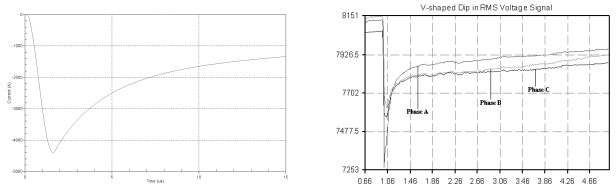
Sag/Dip: A *sag/dip* causes a decrease in between 0.1 pu and 0.9 pu of the RMS voltage or current at the power frequency for durations from 0.5 cycles to one minute. The IEC definition for this phenomenon is *dip*. When not specified otherwise, a 20% sag will be considered an disturbance during which the RMS voltage decreased by 20% to 0.8 pu. Voltage sags are usually associated with system faults but

can also be caused by heavy loads or large motors coming into operation. Fig. 1(a) and (b) show voltage sags that are associated with a motor coming into operation, while Fig. 2(a) and (b) were caused by lightning and transformer energizing respectively.



(a) (b)

Figure 1: (a) Temporary 35% voltage sag caused by motor starting (duration 90 ms) (IEEE (1995)), (b) U-Shaped sag in voltage signal due to Motor start.



(a) (b)

Figure 2: (a) Disturbance in current caused by lightning (IEEE (1995)), (b) V-Shaped sag in voltage signal due to transformer energizing.

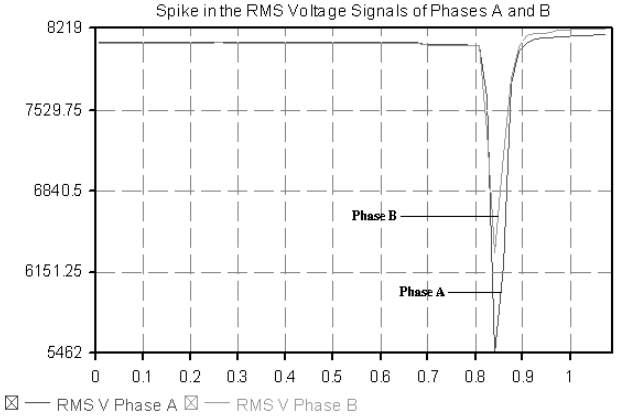
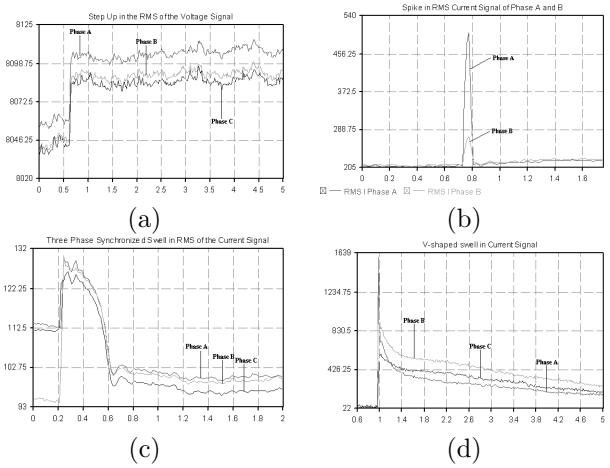


Figure 3: Trough in voltage signal due to transients.

Swell: A swell is defined as an increase to between 1.1 pu and 1.8 pu in RMS voltage or current at the power frequency for durations from 0.5 cycle to one minute. As with sags, swells are usually associated with system fault conditions. However, they are not as common as voltage sags, as they could also be operating in a neighboring circuit. One way that a swell can occur is from the temporary voltage rise on the un-faulted phases during a single line-to-ground (SLG) fault. Swells are characterized by their magnitude (RMS value) and duration. The severity of a voltage swell during a fault condition is a function of the fault location, system impedance, and grounding. A 15% swell, like that shown in the Fig. 4(c), is common on utility feeders.



(a) (b)

(c) (d)

Figure 4: (a) Step-up disturbance in voltage signal, (b) Spike in the current signal due to transient, (c) Synchronous swell in current signal due to motor start, (d) Current transient swell from protective device operation after a fault.

2.3.3 Long duration power disturbance

Increase in the RMS AC voltage (or decrease in RMS AC current) greater than 110% (lesser than 90%) at the power frequency for a duration longer than 1 min. These might be caused by:

- Load going out of operation,
- Capacitor energizing (see Fig. 4(a)), and
- TCUL on.

Decrease in the RMS AC voltage (or increase in RMS AC current) to less than 90% (greater than 110%) at the power frequency for a duration longer than 1 min. The causes often are:

- Load coming into operation,
- Capacitor de-energizing, and
- TCUL off.

3 PROPOSED APPROACH

In this section we introduce the approach that is used in locating the sources from the data. The following are the steps that are involved in building the module (Fig. 5) for such a system:

1. Monitor the power system network and record the data,

2. pre-process the data and extract features,

3. classify the captured current (voltage) signal into predefined categories based on the underlying disturbance that caused them,

4. apply prior knowledge about the source, and

5. decide possible solutions.

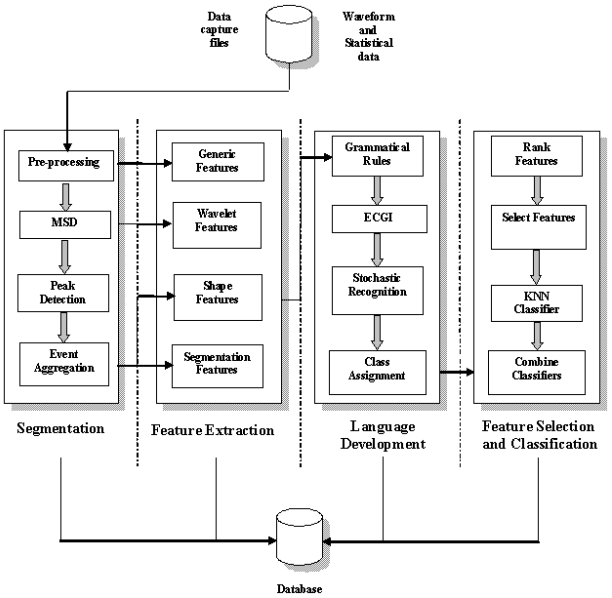


Figure 5: Proposed architecture for detecting and classifying the source of nonstationary disturbances in a distribution system.

3.1 Pre-Processing

3.1.1 High Pass Filtering

The presence of a DC voltage or current in an AC power system is termed DC offset or bias. This phenomenon can occur as the result of a geomagnetic disturbance or due to the effect of half-wave rectification. Power signals often contain DC offsets that limit the capability of the algorithm to estimate and segment the signal. To remove the DC offset, we transmit the raw signal through a high pass digital filter with a cut-off frequency of 20 Hz. This not only removes the DC bias but also the low frequency components that are not representable with the state space model (see Section 3.2.1).

3.1.2 RMS Calculation and Normalization

If the discrete version of the monitored power line signal is represented $\mathbf{z}[n],\ n \in \mathbb{N}^*$, then the RMS of the signal can be calculated as:

$$\mathbf{z_r}[k] = \left(\frac{\sum_{n=(k-1)N_s+1}^{kN_s} \mathbf{z}^2[j+n]}{N_s} \right)^{\frac{1}{2}} \tag{2}$$

where, $N_s = (N \times f)/f_s$ is the number of samples per cycle (256) and N is the total number of samples in the captured signal. The RMS values are updated either at the end of each cycle, or half cycle for reducing the memory for storage or processing.

Once the RMS of the signal is calculated, the signal has to be normalized before extracting the generic features.

$$\mathbf{y}_r[n] = \frac{\mathbf{z}_r[n] - \bar{\mathbf{z}}_r}{s_{\mathbf{z}_r}} \tag{3}$$

where, $(\bar{\mathbf{z}}_r, s_{\mathbf{z}_r})$ are the sample mean and standard deviation of the RMS of the signal respectively.

3.2 Segmentation

Segmentation is the process of dividing an interval of data into sections that are homogeneous with respect to some statistical quality. This is important to not only detect the disturbance but also to localize it within a capture data. There are two methodologies preferred and in the following paragraphs we shall introduce them in detail while also giving valid reasons for choosing one over the other.

3.2.1 Kalman Filter

The Kalman Filtering (Kalman (1960)) provides a means for optimally estimating the voltage and current signal and to track the time-varying parameters. It is a recursive optimal estimator for estimating the 60 Hz voltage and current components. It requires a state variable model for the parameters to be estimated and a measurement equation that relates the discrete measurement to the state variables (parameters). The derivations of the Kalman filter are well described in several excellent references (Barham and Humphries (1969);Brown and Hwang (1992)). The iterative algorithm is given in the Appendix B of this paper.

State Space Model for the current and voltage The mathematical model of the states to be estimated is assumed to be of the form:

$$\mathbf{x}[k+1] = \phi_k \mathbf{x}[k] + \mathbf{w}[k] \tag{4}$$

where, $\mathbf{x}[k+1]$ is the new discrete system state that is to be estimated at a time instant $(k+1)T_s$ and is a linear combination of the previous states and some process noise $\mathbf{w}[k]$ (for allowing variation of the state variables), and $1/T_s$ is the sampling frequency f_s.

The measurement equation is given as:

$$\mathbf{z}[k] = \mathbf{h}_k \mathbf{x}[k] + \boldsymbol{\nu}[k] \tag{5}$$

where, $\boldsymbol{\nu}[k]$ is the random measurement noise, and is assumed to be independent with zero mean and covariance R.

State Variable Representation of a Signal that includes n Harmonics Consider a signal $s(t)$ with a frequency ω, with n harmonics and a magnitude of $A(t)$ defined in the $L^2(R)$ space. These signals are bounded, i.e.

$$\int_{-\infty}^{\infty} |s(t)|^2 dt < \infty \tag{6}$$

where,

$$s(t) = \sum_{i=1}^{n} A_i(t) cos(i\omega t + \theta_i), \tag{7}$$

θ_i is the phase angle of the i^{th} harmonic relative to a reference rotating at $i\omega$.

Since each frequency component required two state variables, in-phase and quadrature-phase components of the harmonics with respect to a rotating reference, the total number of state variables is $2n$. The measurement equation can now be expressed as:

$$\mathbf{z}_k = \mathbf{h}_k \mathbf{x}_k + \boldsymbol{\nu}_k = \begin{bmatrix} cos(\omega k \Delta t) \\ -sin(\omega k \Delta t) \\ \cdots \\ cos(n\omega k \Delta t) \\ -sin(n\omega k \Delta t) \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_{2n-1} \\ x_{2n} \end{bmatrix}_k + \boldsymbol{\nu}_k \tag{8}$$

It should be indicated here that $\mathbf{h}_k$ in this case is a time-varying vector.

Selection of Kalman Filter Parameters To start the Kalman filter recursive estimation, an initial process vector ($\widehat{\mathbf{x}}_0^-$) and the associated initial covariance matrix (P_0^-) are needed. The initial covariance matrix describes, in a statistical sense, the range of variations of the state vector $\mathbf{x}$ from the initial process vector, $\widehat{\mathbf{x}}_0^-$.

1. Initial process vector ($\mathbf{x}_0^-$):
   As the Kalman filter model started with no past measurement, the initial process vector was selected to be zero. The first half cycle is considered to be the "*initialization cycle*".

2. Initial covariance matrix (P_0^-):
   The initial covariance matrix was selected to be a diagonal matrix with the diagonal values equal to 10pu^2.

Once the initialization cycle is complete, the estimated state variables are used as the initial estimate of the state variables for the successive cycle. The Kalman gain vector $\mathbf{k}_k$ and the covariance matrix P_k reach steady-state in about half a cycle. When the above model is used, the steady-state Kalman gain becomes periodic with a period of 1/60 seconds. Its variation include harmonics of 60 Hz. The covariance matrix in the steady-state consists of a constant plus a periodic component. These time variations are due to the time-varying vector in the measurement equation. Thus, after initialization of the model, the Kalman gain vector of the third cycle can be repeated for successive cycles (Girgis et al. (1991)). When all the state variables reach steady state, then the measurement residual ($\mathbf{z}_k - \mathbf{h}_k \widehat{\mathbf{x}}_k^-$) is very small. However, if the signal deviates from the expected behavior, then the residual is very large. This information in combination with a spike detection algorithm (Section 3.2.3) is used to detect nonstationarity.

3.2.2 Multiresolution Signal Decomposition (MSD)

Although Fourier analysis has been used in Power System for analyzing the voltage and current signal, it does not provide good results for *nonstationary* signals. This is because when transforming the complete signal to the frequency domain, the time information gets lost. This means that we can determine all the frequencies present in a signal without knowing where they are present. The MSD overcomes this deficiency by allowing a windowing technique with variable-sized regions, i.e., it allows the use of long time intervals where we want more precise low-frequency information, and shorter intervals where we want high-frequency information.

Since power disturbances are finite energy transient or nonstationary signal, they are decomposed into an approximated and a detailed version of the original signal. Since the detail signal contains sharp edges, transitions, and jumps, the decomposition helps in distinguishing the disturbance from the background signal. However, it is also important to know the level of decomposition and the wavelet suited to detect the slow and the rapid transitions. From experiments it was found that it is sufficient to restrict the decomposition to just four levels for the segmentation process.

If $\mathbf{c}_l$ and the $\mathbf{d}_l$ are the smoothed and detail signals of the original signal at levels l, then

$$\mathbf{c}_l[n] = \sum_k \mathbf{h}[k-2n]\mathbf{c}_{l-1}[k], \tag{9}$$

$$\mathbf{d}_l[n] = \sum_k \mathbf{g}[k-2n]\mathbf{c}_{l-1}[k], \tag{10}$$

where, $\mathbf{h}[n]$ and $\mathbf{g}[n]$ are respectively the associated filer coefficients that decompose the signal into successive levels, and $\mathbf{c}_0[n] = \mathbf{z}[n]$ is the original discrete captured signal.

Choice of the Mother Wavelet The choice of the mother wavelet plays a significant role in detecting and localizing various types of disturbances. Daubechies' was chosen because they are compactly supported with extremal phase and highest number of vanishing moments for a given support width. Also the associated scaling filters are minimum-phase filters. So, from the point of views of fast implementation and varying patterns of the signals, Daubechies' wavelets appeared to be the optimal choice as the mother wavelet for this specific application.

Daubechies' wavelets with 4, 6, 8, and 10 filter order works well in most cases (Santoso et al. (1994)). However,

for sag or over-voltage disturbances, Daubechies' wavelet with four filter coefficients (Daub4) cannot detect or localize the disturbances. From our experiments, it was observed that the fast transients can be detected using the Daub4 or the Daub6 due to their compactness. The slower transient are marked with a smooth amplitude change, and the Daub8 or the Daub10 are long enough to capture them. For the Daub4 wavelet, the scaling function $\phi(x)$ has the form:

$$\phi(x) = c_0\phi(2x) + c_1\phi(2x-1) + c_2\phi(2x-2) + c_3\phi(2x-3) \tag{11}$$

where,

$$\begin{aligned} c_0 &= (1+\sqrt{(}3))/4, \\ c_1 &= (3+\sqrt{(}3))/4, \\ c_2 &= (3-\sqrt{(}3))/4, \\ c_3 &= (1-\sqrt{(}3))/4. \end{aligned} \tag{12}$$

and $\phi(x)$ can be obtained iteratively from (11). The Daub4 wavelet function $\psi(x)$ for the four scaling function is given by:

$$\psi(x) = -c_3\phi(2x) + c_2\phi(2x-1) - c_1\phi(2x-2) + c_0\phi(2x-3). \tag{13}$$

3.2.3 Spike Detection

A *Spike* (or *Trough*) is a data point whose first difference is an outlier value. And outlier is a data point whose first difference is greater than a specified threshold (τ) that can be determined statistically from the time series. A robust estimate of τ should be used so that its value is not affected by the outliers. A spike detection algorithm in combination with the segmentation algorithm determines the location of the disturbance within the signal. If:

$$\mathbf{r}[n] = \mathbf{z}'[n] - \bar{\mathbf{z}'} \tag{14}$$

where, $\mathbf{z}'[n]$ is the first differential of a discrete signal $\mathbf{z}[n]$ that contains the spikes. The output of the spike detection algorithm is:

$$p[n] = \begin{cases} 0, & \mathbf{r}[n] < \tau \\ 1, & \mathbf{r}[n] \geq \tau \end{cases} \tag{15}$$

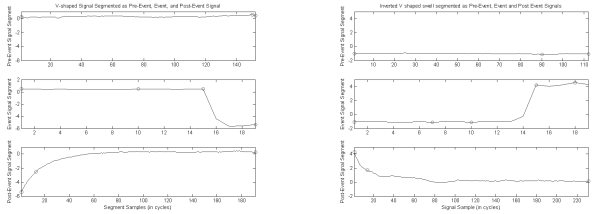
where, $\tau = 10 * \sigma_r$, and

$$\sigma_r^2 = \frac{1}{N}\sum_{i=1}^{N}\left(r[i] - \mu_r\right)^2. \tag{16}$$

The morphological filters of (Ji et al. (2008);Lu et al. (2006)) could also be used for the spike detection. Fig. 6 shows the result of segmenting the voltage and current signal due to an incoming load into the pre-disturbance, disturbance and post-disturbance categories.

4 AUTOMATIC FEATURE EXTRACTION AND SELECTION

In this section, we discuss the automatic feature extraction and the optimal feature selection module.



(a) (b)

Figure 6: (a) V-shaped sag segmented into three subcategories, (b) V-shaped swell segmented into three subcategories (knots are shown as circles).

4.1 Feature Extraction

It is the process of transforming the classification data into a format that highlights the class differences in the data and also converts it to a form compatible with the classifier. We use the heuristics guidelines of compactness, bounds and clustering ability from (Crossman et al. (2003)) for extracting the signal features. Accordingly, we found the following feature families that are sufficient to characterize the disturbance behaviors.

4.1.1 Generic Features

The basic features that are extracted from the normalized RMS of the signal or the segmented signal are:

1. Minimum: $n_{\mathbf{y}_r} = min(\mathbf{y}_r[i])$, $i \in [a,b]$
2. Maximum: $m_{\mathbf{y}_r} = max(\mathbf{y}_r[i])$, $i \in [a,b]$
3. Range: $r_{\mathbf{y}_r} = m_{\mathbf{y}_r}$-$n_{\mathbf{y}_r}$
4. Length (duration): $l_{\mathbf{y}_r} = b - a$
5. Average: $a_{\mathbf{z}_r} = \frac{\sum_{i\in[a,b]} \mathbf{z}_r[i]}{b-a}$
6. Fluctuation: $f_{\mathbf{z}_r} = \left(\frac{\sum_{i\in[a,b]} (z_r[i]-\bar{\mathbf{z}}_r)^2}{b-a}\right)^{\frac{1}{2}}$

where, $\mathbf{y}_r$ is a given length segment (3), and i is the signal segment sample index. The minimum, maximum, range. and average give the statistical nature of the signal and they are self-descriptive. The *fluctuation index*, gives the amount of variations in each of the segment. The other generic features distinguish between identically shaped signals and are useful for the advanced classification process. Table 1 gives the generic feature vector extracted for a TCUL on operation.

4.1.2 Harmonic Features

Certain disturbances cause harmonic distortion in the voltage and the current signal. *Harmonic Distortion* is mainly caused by the operation of nonlinear loads and devices in the power system. *Total Harmonic Distortion (THD)*

FEATURES	VALUES
Number of Spikes	0
Fluctuation Index	2.4147
RMS Signal Maximum (p.u.)	1.0928
RMS Signal Minimum (p.u.)	0.89907
Number of inflections	1
RMS Signal Range	0.1937
Transient Detected (1-Yes, 0-No)	0

Table 1: Generic feature vector for a TCUL.

(Sabin et al. (1999)) can describe the amount of harmonic distortion by normalizing the total voltage harmonic component content by the fundamental frequency component. Accordingly,

$$THD = \frac{1}{\mathbf{z}_{r,1}}(\sum_{k=2}^{\infty} \mathbf{z}_{r,k}^2)^{\frac{1}{2}} \times 100\% \tag{17}$$

where, $\mathbf{z}_{r,k}$ is the RMS value of the k^{th} harmonic and $\mathbf{z}_{r,1}$ is the RMS value of the fundamental.

4.1.3 Segmentation Features

For recognition of geometrically shaped signals, localization of the disturbance is of primary importance. The features that are extracted based on just segmentation of the signal are: location in the captured data, number of disturbances, individual magnitude, separation between multiple disturbances, and number of phases involved (synchronous or non-synchronous). If there are multiple disturbances that are related to each other, then they should be temporally aggregated to identify or confirm the disturbance. If they are unrelated, they should be each treated differently and the disturbance that is most likely to cause alarm should be used to label the capture data. It is very difficult to specify when the disturbances in a capture data are related. Usually if the time separation between them is more than 1 minute, they should be treated as two separate disturbances. Also it was a general observation that most disturbances that are synchronous in all the phases are most likely to be normal power system disturbances (load related) and most abnormal disturbances are asynchronous in nature. Table 2 gives the generic feature vector extracted for the Step-up disturbance.

FEATURES	VALUES
Change Direction (Up: 1, Down: -1, Plateau: 0)	1
Number of disturbances	1
Disturbance Separation (Max, Mean, Min)	0,0,0
Phase Information (Exists-1, N/A-0)	[1 1 1]

Table 2: Example segmentation feature vector for a step-up disturbance.

4.1.4 Shape-based features

Concavity/Convexity This is a rough estimate of the curvature of a discrete signal over a particular segment. Given a function $f : \mathbb{R} \mapsto \mathbb{R}$, is *concave* $\iff$

$$f(t) \geq \frac{pf(t-q) + qf(t+p)}{p+q}, \forall p, q > 0, \forall t > q. \tag{18}$$

A function is f *convex* if $(-f)$ is *concave*. For a discrete signal, the curvature is a measure of deviation from a straight line or closeness to a circle. Both indices are calculated by fitting a straight line curve between the two endpoint data samples of a signal segment, and then calculating the ratio of the number of samples in the segment that are above (concavity) and below (convexity) this line.

Let the equation of the line be given by $\mathbf{f} = m\mathbf{y}_r + b$, then from the definition of concavity and convexity:

$$c_v = \frac{\sum\limits_{i\in[a,b]} \begin{cases} 1 & \text{if } \mathbf{y}_r[i] - \mathbf{f}[i] > 0 \\ 0 & \text{if } \mathbf{y}_r[i] - \mathbf{f}[i] \leq 0 \end{cases}}{b-a} \tag{19}$$

$$c_x = \frac{\sum\limits_{i\in[a,b]} \begin{cases} 1 & \text{if } \mathbf{y}_r[i] - \mathbf{f}[i] < 0 \\ 0 & \text{if } \mathbf{y}_r[i] - \mathbf{f}[i] \geq 0 \end{cases}}{b-a}. \tag{20}$$

This feature gives a basic idea of the segment's shape and it was found to be very resistant to noise.

Signal Bends The bends in a signal are regions in the signal where the curvature of the signal changes. The shape of the signal can be approximated by knowing the number of times the RMS of the signal changes its direction. The number of bends in the signal is calculated by detecting the number of times the line $\mathbf{f} = \frac{m_{\mathbf{y}_r} + n_{\mathbf{y}_r}}{2}$ intersects the signal $\mathbf{y}_r[n]$.

4.1.5 Detecting High Frequency Transients

High Frequency Transients in a signal could be detected by a combination of a *higher order difference filter* and a spike detection algorithm (Section 3.2.3). The interested readers can refer to (Pankajakshan and Kumar (2007)) for more details on the higher-order difference filter.

4.1.6 Wavelet and Wavelet Packet Features

Detail Wavelet Features We perform the discrete wavelet decomposition on the signal (DWT) to the twelfth level of resolution (9) (10). From the wavelet coefficients, we compute the average energy content of the coefficients at each resolution. Features are then extracted from the 13 sub-bands (12 wavelet sub-bands and one approximation sub-band).

The i^{th} element of a feature vector is given by

$$\mathbf{v}_i^{dwt} = \frac{1}{n_i}\sum_{j=1}^{n_i} W_{i,j}^2, \;\; i = 1, 2, \ldots, 13, \tag{21}$$

where, $n_1 = 2^{11}, n_2 = 2^{10}, \ldots, n_{12} = 2^0, n_{13} = 2^0$; $\mathbf{v}_i^{dwt}$ is the i^{th} feature element in a DWT feature vector; n_i is the number of samples in an individual sub-band; and $W_{i,j}^2$ is the j^{th} coefficient of the i^{th} sub-band. As a result, a DWT feature vector is formed as

$$\mathbf{v}^{dwt} = [v_1^{dwt}, v_2^{dwt}, \ldots, v_{13}^{dwt}]^T \tag{22}$$

In the following few paragraphs, we introduce additional features that are a useful measure of shapes and noise levels.

- *Wavelet Coefficient Average*(WA_l): This feature is the average value of detail wavelet coefficients at level l over the segment. When calculated for the Daub1 mother wavelet, WA_l gives an indication of the slope of the signal.

$$\mathrm{WA}_l = \frac{\sum\limits_{i\in[b,a]} \mathbf{d}_{l_i}}{b-a} \tag{23}$$

  where, $\mathbf{d}_l$ is the l^{th} detail level of wavelet coefficients; $\mathbf{d}_{l_i}$ is the i^{th} coefficient in the l^{th} level.

- *Wavelet Coefficient Energy (WE$_l$)*: WE_l gives the energy of the detail wavelet coefficients at level l over the segment. For normally plateau segments, WE_l can be a good indication of the noise level and frequency over the segment. The noise distorts the 4^{th} level wavelet coefficient, but not the 1^{st} level wavelet coefficients.

$$\mathrm{WE}_l = \frac{\sum\limits_{i\in[b,a]} \mathbf{d}_{l_i}^2}{b-a}. \tag{24}$$

- *Wavelet Quarters WQ:* This feature consists of four values per detail coefficient and is calculated as

$$\begin{aligned}
\mathrm{WQ}_l^1 &= \frac{\sum\limits_{i\in[a,m]} \mathbf{d}_{l_i}}{m-a}, \quad \textit{if } \mathbf{d}_{l_i} > 0 \\
\mathrm{WQ}_l^2 &= \frac{\sum\limits_{i\in[a,m]} \mathbf{d}_{l_i}}{m-a}, \quad \textit{if } \mathbf{d}_{l_i} < 0 \\
\mathrm{WQ}_l^3 &= \frac{\sum\limits_{i\in[m,b]} \mathbf{d}_{l_i}}{b-m}, \quad \textit{if } \mathbf{d}_{l_i} > 0 \\
\mathrm{WQ}_l^4 &= \frac{\sum\limits_{i\in[m,b]} \mathbf{d}_{l_i}}{b-m}, \quad \textit{if } \mathbf{d}_{l_i} < 0
\end{aligned} \tag{25}$$

  where, $m = \frac{(b-a)}{2}$. Upon careful inspection it can be seen that the (26) defines the average wavelet coefficient value for each of the four quadrants of the segment. By name, the four quadrants are: *begin-positive, begin-negative, end-positive,* and *end-negative* respectively. By separating the segment in half, the behavior of the signal entering into and exiting a segment can be identified. By averaging positive and negative coefficients separately, we can avoid giving signals with high fluctuation the same feature values of a signal with very low fluctuation.

Wavelet Packet Features We perform the wavelet packet (Mallat (1999)) multiresolution analysis to the fifth level of resolution to obtain 32 sub-bands. Each sub-band contains a total of 128 wavelet packet coefficients. From each sub-band at the fifth level of resolution, we compute the average energy content in the packet coefficients as:

$$\mathbf{v}_i^{wp} = \frac{1}{n_i}\sum_{j=1}^{n_i} p_{i,j}^2, \;\; i = 1, 2, \ldots, 32 \, and \, n_i = 128 \, \forall i, \tag{26}$$

where, $\mathbf{v}_i^{wp}$ is the i^{th} feature in a wavelet packet feature vector, n_i is the number of sample in each sub-band, and $p_{i,j}$ is the j^{th} wavelet packet coefficient in the i^{th} sub-band. The WP feature vector is represented as:

$$\mathbf{v}^{wp} = [v_1^{wp}, v_2^{wp}, \ldots, v_{32}^{wp}]^T \tag{27}$$

4.2 Optimal feature subset selection and ranking

After extracting the features, sometimes there could be some of them that are redundant. The problem of optimal *feature selection* is defined as follows:

Definition 1. *Given a set of* d *candidate features, select a subset of size* m *($m \leq d$) that leads to the smallest classification error (P_ϵ).*

This procedure can reduce not only the cost of recognition by reducing the number of features that need to be collected, but in some cases it can also provide a better classification accuracy due to finite sample size effects.

Let $\mathcal{F}_d \in \mathbb{R}^{n\times d}$ be the given set of features with cardinality d and let m represent the cardinality of the desired subset $\mathcal{F}_m \in \mathbb{R}^{n\times m}$, $\mathcal{F}_m \subset \mathcal{F}_d$.

$$\mathcal{F}_m = A_m^T \mathcal{F}_d \tag{28}$$

with $A_m^T A_m = I_m$, and I_m is identity. This can be viewed as a linear transformation of $\mathcal{F}_d$ using a transformation matrix:

$$A_m = \begin{bmatrix} I_m \\ [\mathbf{0}]_{(d-m)\times m} \end{bmatrix} \tag{29}$$

or any matrix that is a permutation of the rows of A_m. Without loss of generality, we consider the transformation matrix A_m and rewrite the corresponding covariance matrix of $\mathcal{F}_d$ as:

$$\Sigma = \begin{bmatrix} \{\Sigma_{11}\}_{m\times m} & \{\Sigma_{12}\}_{m\times(d-m)} \\ \{\Sigma_{21}\}_{(d-m)\times d} & \{\Sigma_{22}\}_{(d-m)\times(d-m)} \end{bmatrix} \tag{30}$$

The feature ranking criterion can be identified by an evaluation function $J : \mathbb{R}^{n\times d} \mapsto \mathbb{R}$. Higher the value of $J(\mathcal{F}_d)$, better is the set $\mathcal{F}_d$. Following the above consideration, if P_ϵ (Appendix A) is the classification error, a natural choice for the evaluation function is:

$$J = (1 - P_\epsilon) \tag{31}$$

If the number of known classes is C (see Table 3) for a d dimensional problem with mean μ_i, covariance matrix Σ_i and a priori probability of Π_i (Table 4) of class

i $(i = 1, 2, \ldots, C)$. Then let $\mu_0 = \sum_{i=1}^{C} \Pi_i \mu_i$. $\Sigma_B = \sum_{i=1}^{C} \Pi_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T$, $\Sigma_W = \sum_{i=1}^{C} \Pi_i \Sigma_i$, and $\Sigma_T = \Sigma_B + \Sigma_W$ are between-class, within-class and total covariance matrices respectively.

Disturbance	Assigned Labels
Step Down	1
Step Up	2
Trough	3
Spike	4
V-shaped Sag	5
U-shaped Sag	6
V-shaped Swell	7
U-shaped Swell	8

Table 3: Disturbances and their assigned class labels.

Disturbance Class	Prior Probability
Step Down (1)	0.19174
Step Up (2)	0.15044
Trough (3)	0.11209
Spike (4)	0.064897
V-shaped sag (5)	0.19469
U-shaped sag (6)	0.10029
V-shaped swell (7)	0.047198
U-shaped swell (8)	0.13864

Table 4: Empirically calculated prior probabilities for each disturbance.

5 CLASSIFICATION

Approaches to designing a classifier may be divided into two principal categories:

1. the decision-theoretic approach, and
2. the syntactic approach.

The decision-theoretical approach relies on the utilization of decision functions for classifying the pattern vectors. It is ideally suited for applications where the patterns can be meaningfully represented in vector form (like the features extracted earlier).

5.1 k-nearest neighbor (kNN) algorithm

The method of Nearest Neighbor, proposed by T. M. Cover and P. E. Hart (Cover and Hart (1967);Wagner (1971)), is a very efficient nonparametric approach to classification. This decision rule assigns an unclassified sample point to the nearest of a set of previously classified points based on the estimated density $P(\mathbf{x})$. This rule is independent of the underlying joint distribution on the sample points and their classifications.

$$P(\mathbf{x}) \cong \frac{k}{NV} \tag{32}$$

where, V is the volume surrounding x, N is the number of samples, k is the number of samples inside V. The unconditional density is estimated as,

$$P(\mathbf{x}|\omega_i) = \frac{k_i}{N_i V} \tag{33}$$

And the priors can be estimated as,

$$\Pi(\omega_i) = \frac{N_i}{N} \tag{34}$$

The posterior probability then becomes,

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)\Pi(\omega_i)}{P(\mathbf{x})} = \frac{\frac{k_i}{N_i V} \cdot \frac{N_i}{N}}{\frac{k}{NV}} = \frac{k_i}{k} \tag{35}$$

The discriminant function thus becomes,

$$g_i(\mathbf{x}) = \frac{k_i}{k} \tag{36}$$

For a given unlabeled example $\mathbf{x}_u \in \mathbb{R}^n$, find the k closest labeled examples in the training data set and assign $\mathbf{x}_u$ to the class that appears most frequently within the k-subset.

5.2 Stochastic recognition

5.2.1 Signal Morphology Description

The *syntactic* approach to pattern recognition possesses the structure-handling capability lacked by the decision theoretic approach. It is important that the characteristics of the signal or patterns are best described by structural relationships. Artificial Neural Networks (ANN) based shape discriminators described in (Angrisani et al. (1998)) have been used in the past to distinguish transient shape. However, these approaches are often tuned to the templates that are used in the training process. Any geometric translations, scaling, stretching or others will affect the performance of the recognition algorithm.

5.2.2 RMS Signal Structural Sampling

Each segment of the RMS current or voltage can be sampled in terms of some pre-defined sub-waveform *primitives*. The elementary properties used to describe the hierarchical structure of an object are called as *primitives*. The core library of primitives includes the following:

- Simple-plateau and quantized-plateau, used to recognize when a signal is constant, though perhaps noisy.
- Trending-up and trending-down, used to recognize when the signal is periodically reaching new highs (or new lows).

Each of these primitives is parameterized to help deal with noisy data and they are assigned a symbol that forms part of a relational structure representing a pattern. We label them with the lower case alphabets $c - r$ and $t - x$ (Table 5). Each of them also have a minimum-duration parameter that helps in avoiding trivially short segments. We illustrate some example patterns and their associated words in Table 6.

PRIMITIVES	DESCRIPTION
↑	Very Fast Trending Rise (c-g)
↗	Fast Trending Rise (h-l)
→	Plateau (m)
↘	Fast Trending Fall (n-r)
↓	Very Fast Trending Fall (t-x)

Table 5: Example of terminals and their pattern primitive interpretations.

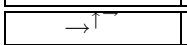
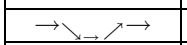
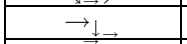
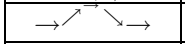
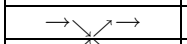
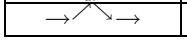
PATTERNS	STRING REPRESENTATIONS
	mcm
	momkm
	mxm
	mkmom
	mrhm
	mhrm

Table 6: Example patterns and their representation as strings.

5.2.3 Assigning Primitives to Segments

It is necessary that the piecewise linear signal segments, generated by the segmentation process are assigned to the right primitives. The task of primitive assignment is not an easy one because the assignment should be insensitive to noise or short term variations in the segments, and it should not be restricted by the length of the segment. The recursive partitioning of the segment procedure may slice one large segment into a few smaller segments. Consequently, some adjacent segments may have the same state label. A simple search through the list of segments examines all adjacent pairs. If a pair of adjacent segments has the same state, the two segments are combined into one. Also if a segment labeled as a steady state is too short to be significant, it is merged to its neighboring segment. The subsequences thus obtained is compared with some extracted primitive templates from some test signals. Since there is a database of standard subsequence templates, we use a *fast similarity search* (Agrawal et al. (1995)) to choose the closest one.

5.2.4 Grammar Inference and Parsing

The *grammar* is a set of rules of syntax for the generation of *sentences* from the given *primitives* and decides the ones that are permissible. The grammar normally divides the string into the grammatically correct ones and the others. In case of ambiguity, it is necessary to resolve by approximation, stochastic approach, similarity and error-correcting parsing or transformation. The distance measure is defined between a sentence representing a known pattern and a sentence representing an unknown pattern. The distance is measured in terms of number of errors after insertion, deletion or substitution. *Grammatical Inference* aims at learning models of languages from examples of sentences of these languages. It is concerned mainly with the procedures that can be used to infer syntactic or production rules of an unknown grammar G based on a finite set of strings I from $L(G)$, the language generated by G and possibly also on a finite set of strings from the complement of $L(G)$ (see Appendix C).

Stochastic Grammar A *stochastic* system is one in which each step has more than one production rules for deriving it and probabilities are assigned to alternatives in order to indicate likelihood of occurrence or to guide the process of selection.

The Error-correcting Grammatical Inference (ECGI) (Rulot and Vidal (1987)), is a GI heuristic similar to a Viterbi-like dynamic algorithm (Amengual and Vidal (1998);Forney (1973)) that was explicitly designed to capture relevant regularities of concatenation and length exhibited by substructures of uni-dimensional patterns. It relies on error-correcting parsing to build up a stochastic regular grammar through a single incremental pass over a positive training set I_+. As each new string is presented, the learning procedure attempts its recognition through the current automaton. Initially, a trivial grammar is built from the first string of I_+. Then, for every new string that cannot be parsed with the current grammar, a standard *error-correcting* scheme is adopted to determine a string in the language of the current grammar that is error-correcting closest to the input string. From these results the current grammar is updated by adding a number of rules and (non-terminals) that permit the new string, along with other adequate generalizations to be accepted. Similarly, the parsing results are also used to update frequency counts from which probabilities of both non-error and error rules are estimated.

The ECGI Inference Method Let $G = (V, \Sigma, P, S)$ be a regular grammar. The error rules of G can be defined as follows (Chodorowski and Miclet (1998)):

Definition 2. For each rule in P, each $a, b \in \Sigma$, define the error rules associated to P by:
Insertion (of a):
$A \to aA, \forall (A \to bB) \in P,\ A \to ab, \forall (A \to b) \in P$;
Substitution (of a by b):
$A \to aB, \forall (A \to bB) \in P,\ A \to a, \forall (A \to b) \in P$;
Deletion (of b):
$A \to B, \forall (A \to bB) \in P,\ A \to \in, \forall (A \to b) \in P$;

The extended grammar $G' = (V, \Sigma, P', S)$ is obtained by adding the error rules to P

Definition 3. The optimal corrective derivation of $x \in \Sigma^*$ is the derivation $D_{G'}(x)$ of x using the minimal number of error rules in P'.

The inner loop of the following procedure computes the optimal corrective derivation of a sentence with respect to a grammar. The whole procedure infers a grammar which depends on the order of the examples.

ECGI Algorithm

input I_+
output *Grammar* G *accepting* I_+

begin
$a \leftarrow I_+^1(= a_1 \dots a_i \dots a_n)$;
$V \leftarrow A_0, \dots, A_n, -1$;
$\Sigma \leftarrow a_1 \dots a_i \dots a_n$;
$P \leftarrow (A_{i-1} \rightarrow a_i A_i), i = 1, \dots, n \cup A_{n-1} \rightarrow a_n$;
$S \leftarrow A_0$;
$G \leftarrow (V, \Sigma, P, S)$;*//current grammar*
for $i \leftarrow 2$ ***until*** $|I_+|$*//taking a new example*
$b \leftarrow I_+^i$;
$D \leftarrow D_G(b)$;
$G \leftarrow G \cup D$;*// Adding the corrective rules to the current grammar*
end for
return *G;*
end *ECGI*

The Maximum Likelihood that a string $x \in V^*$ be generated by the stochastic (error-correcting) extended grammar of G, G', is given by:

$$P_{G'}(x) = max_{\forall D'_G(x)} \Pi_{\forall r_i \in D'_G(x)} P(r_i) \qquad (37)$$

where, $D_{G'}(x)$ is any error-correcting derivation of x from G, and $P(r_i)$ is the probability of the i^{th} rule of $D'_G(x)$. For a C-class problem, with each class C_i represented by a grammar G_i, the generation probabilities can be utilized for classification in the usual way:

$$x \in C_i \text{ iff } P_{G'_i}(x)\Pi_i > P_{G'_j}(x)\Pi_j \; i \neq j;$$
$$j = 1 \dots C, \forall x \in V^*$$

where, Π_j, $j = 1 \dots C$ is the a priori probability of class C_j.

Disturbance Class	Number of rules generated
1	51
2	56
3	123
4	121
5	117
6	92
7	75
8	115
Total unique rules	368

Table 7: Rule generation results for each class.

6 RESULTS

The segmentation approach that was proposed in Section 3.2 was first tested on the simulated data and then on the acquired data. The following are the results from these experiments.

Rules Executed	Rule Probability	Rule Frequency
S→mM	0.030857	630
M→mM	0.09977	2037
M→rR	0.012734	260
R→mM	0.011559	236
M→mM	0.09977	2037
M→m	0.015575	318

Table 8: Example rules for a Class 1 disturbance.

6.1 Experiments on simulated data

A single phase model was built using Simulink® for simulating the motor start voltage disturbance (Kezunovic et al. (1998)). The model is as shown in Fig. 7 and the sag in the voltage signal due to the load is shown in Fig. 8. The MSD

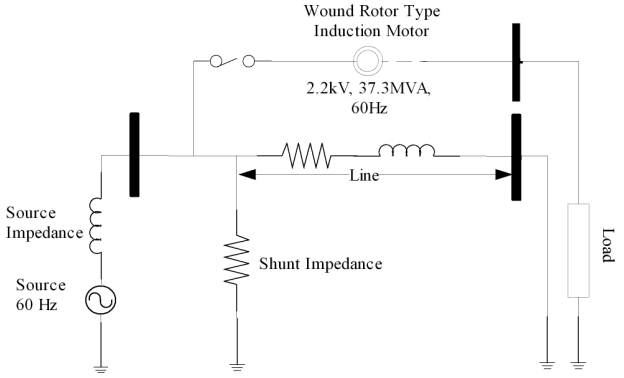


Figure 7: Single Phase model for simulating motor start disturbances.

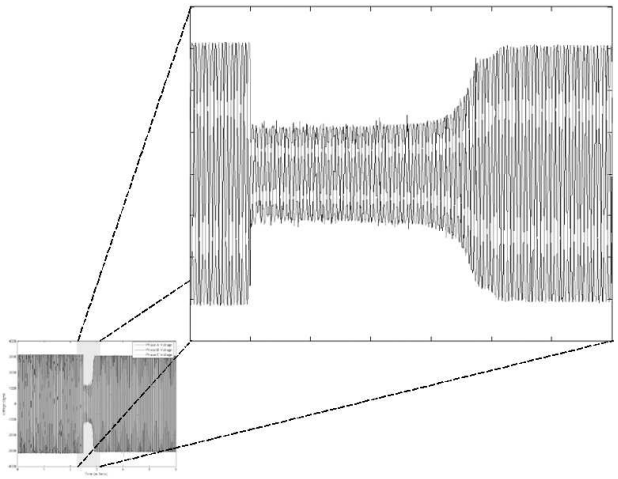
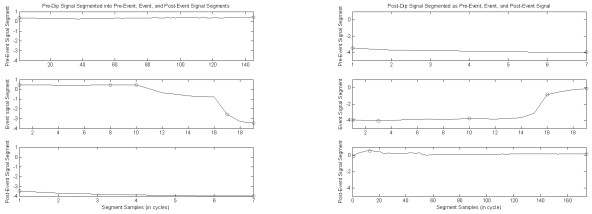


Figure 8: Disturbance in Voltage Signal during Motor Start.

segmentation algorithm was done on one such simulation of a motor coming into operation and the 6 sub-categories are as shown in Fig. 9.

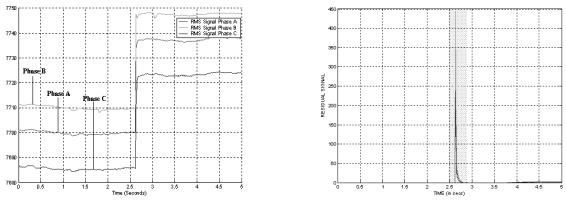


(a) (b)

Figure 9: Voltage signal due to a motor starting operation is segmented into six sub-categories (knots are shown as circle) (a) Pre-disturbance, (b) Post-disturbance voltage signals.

6.2 Experiments on recorded data

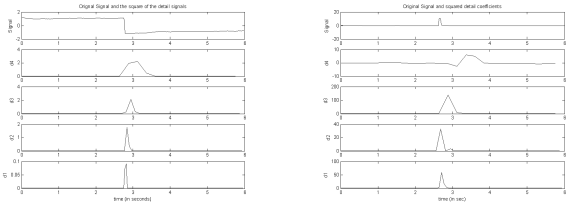
The Fig. 10(a) shows an example of the RMS of the voltage signal for a step-up disturbance and the the Fig. 10(b) shows the stepping disturbance segmented using the Kalman Filter and the output of the spike detection algorithm. The Fig. 11-12 gives the detection and localiza-



(a) (b)

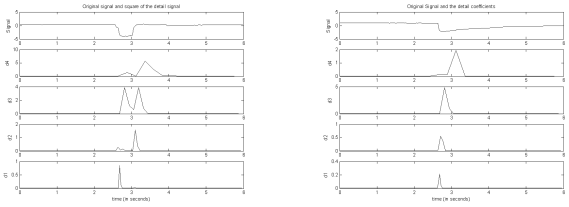
Figure 10: Using the residual sequence for detecting transition points (a) Example of step up measurement data and, (b) disturbance segmented using the Kalman Filter residual sequence.

tion of the disturbances on different captured power system signals. The disturbances shown in Fig. 11(a) and Fig. 11(b) are fast transients and can be detected using the level 1 detail coefficients. However, the disturbance shown in Fig. 12(a) and Fig. 12(b), the recovery of the signal are slow and can be detected only using the Daub6 or the Daub8.



(a) (b)

Figure 11: MSD on (a) step down disturbance signal, (b) spike swell disturbance signal.



(a) (b)

Figure 12: MSD on (a) U-shaped sag disturbance signal, (b) V-shaped sag disturbance signal.

6.3 Validating the Classifier

The results of the classification is as given in Table 9. We obtain a classification accuracy of about 96%. From the Confusion Matrix in Table 10, it can be observed that the algorithm sometimes fails to distinguish between an impulse trough transient and incoming load voltage variation. This is because of the similarity in shape between the two disturbances. Additional knowledge on either could help in clearly differentiating them.

Correctly Classified Test Data	331 (95.389%)
Incorrectly Classified Test Data	16 (4.611%)
Total Number of Test Data	347
Mean absolute error	0.014
Root Mean Squared Error	0.1068
Relative absolute error	6.5172%

Table 9: Classification results.

There is no definite function that evaluates the tolerance between misclassifying a disturbance or rejecting it as unknown. For example, the cost of classifying an overcurrent disturbance (abnormal, possibly an emerging condition) as motor start (normal)-FALSE NEGATIVE is higher than the cost of classifying a motor start as an overcurrent disturbance-FALSE POSITIVE. In evaluating the relative performance of the various methods, the research team uses the following general guidelines:

1. The method must produce a high percentage of "TRUE POSITIVES," i.e., outputs that correctly classify the data captures that the method is designed to recognize.

2. The method must produce very few "FALSE POSITIVES," i.e., outputs that indicate a data capture has one cause when the real cause was something else.

3. Minimization of false positives is much more important than maximization of true positives. In other words, it is more acceptable to have data captures that a method cannot classify that it is to have even a few data captures that it classifies incorrectly.

Table 10: Confusion matrix for the classifier

PREDICTED\ACTUAL	1	2	3	4	5	6	7	8
1	64	0	0	0	0	0	0	0
2	0	51	0	0	0	0	0	0
3	0	0	43	0	6	0	0	0
4	0	0	0	19	0	0	3	0
5	0	0	0	1	59	0	0	0
6	0	0	0	0	2	33	0	0
7	0	0	0	0	0	0	15	2
8	0	0	0	0	0	0	2	47

Class Labels	True Positive	False Positive	Precision
1	1	0	1
2	1	0	1
3	0.878	0	1
4	0.864	0.003	0.95
5	0.983	0.028	0.881
6	0.943	0	1
7	0.882	0.015	0.75
8	0.959	0.007	0.959

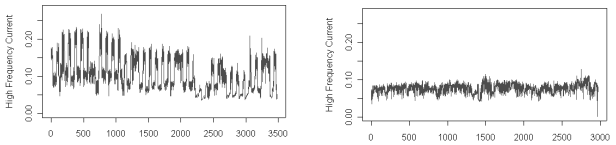
Table 11: Classification accuracy by disturbance.

7 TEMPORAL AGGREGATION FOR DETECTING EXISTING CONDITIONS

7.1 Case study

Electric power distribution systems experience faults for a variety of reasons. The majority of the distribution faults are due to degradation of distribution devices because of natural causes. When these devices begin to deteriorate, intermittent emerging faults persist in the system from as little as several days to several months. Hence, characterizing device failure behavior is essential to allow scheduling of distribution maintenance and execute remedial actions aimed at replacing the failing equipment prior to the occurrence of faults and service interruptions. Since an exact failure process of a piece of equipment or system is not completely known, we usually monitor a large number of parameters to relate the behavior of the parameters to the faults. In (Kim et al. (1999)), it was demonstrated that emerging fault conditions provide detectable signals, thereby validating the concept of using on-line monitoring conditions to detect them. Even though there are a variety of conditions that can lead to supply interruptions during fair weather, there are certain emerging events that produce measurable electrical activity in advance of a failure or outage. Dynamic response to these instabilities may involve large deviations from normal conditions or instability. Many characteristics may not be known with absolute certainty. The basic issue to be addressed is there are two forms of uncertainty: small deviations from observed phenomena and completely unseen phenomena. The latter issue is almost impossible to be recognized automatically but the former could be handled using a timed-event trend analysis.

It was shown in (Manivannan and Pankajakshan (2003)) that capacitor banks in a distribution feeder causes masking of high frequency transients. Conversely, changes in the observed daily cycle of the high frequency data can be useful to obtain future forecasts for the data. By comparing the forecast data with the actual data, it is possible to know if there is any deviation from the normal behavior. In Fig. 13(a) we see a noticeable trend in the high frequency current (obtained by using a high pass filter on the current data) during the month of December and absent after one month (see Fig. 13(b)). The data is scaled to highlight the difference. A simple Correlogram plot as shown in Fig. 14



(a) (b)

Figure 13: High frequency currents in the month of (a) December, and (b) January.

in combination with the classification algorithm helped in identifying the trends as daily cycles of a capacitor coming into and going out of operation. Once the cause has been identified, the change or absence in the trends should be a cause for alarm as it simply means the non-operation of that particular device. The necessity for a temporal aggregation (Kim (2000);Kim et al. (2004)) is highlighted by this simple example. By using the first sequence of data, one can predict cyclic variations in the subsequent sequences. The residual sequence obtained after comparison of the actual data with the predicted data can give us an indication of the level of deviation from the normal behavior. By using a simple Autoregressive-Moving Average (ARMA) model, we can predict the successive states from the previous states. The model is updated or recalculated every time a deviation from the predicted behavior is observed.

If we assume that $\mathbf{x}$ is the original RMS of the high

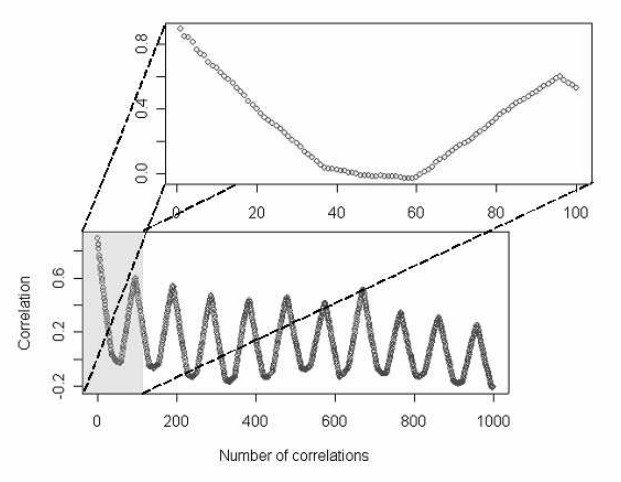


Figure 14: Correlogram plot for lag 1000 and enlarged version to show cycle length.

frequency current, then the series can be modeled as:

$$(1-l)(1-l^n)\sum_{m=0}^{M-1}\alpha_m \mathbf{x}[j-m] = \sum_{k=0}^{K-1}\beta_k \in [j-k] \quad (38)$$

In this specific case, n the lag was calculated to be about 100, $M = 16$ and $K = 12$. The plot of the predicted and actual series with the bounds is as shown in Fig. 15.

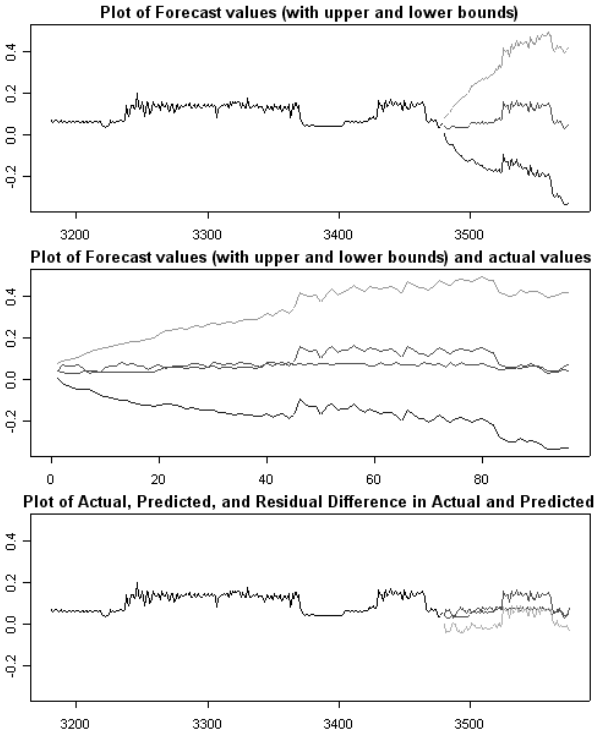


Figure 15: Plot of actual, predicted, upper bound, and lower bound for the High Frequency current data.

8 CONCLUSIONS

In this paper we presented a unified framework for automatically classifying disturbances based on their underlying cause. In this process, we used a combination of a decision-theoretic and syntactic approach. The classification method starts with segmenting the current and voltage signal into the stationary and nonstationary segments. We characterize the segments of each phase and also compare the corresponding segments between the different phases. Syntactic pattern representation of the RMS of a power signal is used to provide the necessary level of abstraction between the raw data and the decision making system. It is better than template matching techniques especially when the input signal is noisy, scaled or warped in comparison to the templates. Eight types of disturbance sources were considered here and the testing was done on measurements acquired from a distribution network with a classification accuracy of more than 95%. The error in classification is mainly due to the similarity in shapes between the disturbances. A more direct way of distinguishing them would be by using temporal aggregation. A case study was provided to highlight the use of temporal aggregation in detecting failing or failed devices.

It is also possible to develop the system so that it can automatically build the templates in limited cases by looking at a set of training data (Miclet and Quinqueton (1988)). For example, in the case of a capacitor switching operation, by determining the *reactive power* changes or the presence of transients and harmonics (Girgis et al. (1993)), it is possible to identify unsuccessful energizing, blown fuses, or even failed capacitor banks. Future work could be based on extending this framework to include abnormal disturbances that affect the distribution system immediately.

A DISTANCE METRICS

Formally a metric is a function $d(\cdot,\cdot)$ from $\mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ that defines the degree of closeness between two vectors $\mathbf{x}$, $\mathbf{y}$ and satisfies the properties of non-negativity, definiteness, symmetry, triangle inequality and identity.

Definition 4. *The statistical distance or Mahalanobis distance between two vectors* $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_p)^T$ *in the p-dimensional space* $\mathbb{R}^p$ *is defined as*

$$d(x,y) = ((x-y)^T\Sigma^{-1}(x-y))^{\frac{1}{2}} \quad (39)$$

and $d(x,0) = (x^T\Sigma^{-1}x)^{\frac{1}{2}}$ *is the norm of* x.

Let $\delta(i,j)^2 = (\mu_i - \mu_j)^T\Sigma_W^{-1}(\mu_i - \mu_j)\ \forall i \neq j, i,j = 1,2,\ldots,C$ be the squared Mahalanobis distance between the classes i and j.

Lemma A.1. *Given the normally distributed data with means* $\mu_1, \ldots, \mu_C$*, equal covariance matrices* $\Sigma_1 = \ldots = \Sigma_C$*, and equal a priori probabilities* $\pi_1 = \ldots = \pi_C$ *the*

error rate P_ϵ is bounded by

$$P_\epsilon \leq \frac{C-1}{C} \sum_{i=1}^{C} \Phi(-\frac{1}{2} \min_{j=1,\ldots,C, j\neq i} \delta(i,j)) \qquad (40)$$

where, Φ is the (cumulative) distribution function of the standard normal distribution. The proof of this lemma is given in the appendix of (Luebke and Weihs (2005))

B KALMAN FILTERING

The observation (measurement) of the process is assumed to occur at discrete points in time in the form (8):

$$\mathbf{z}_k = \mathbf{h}_k \mathbf{x}_k + \boldsymbol{\nu}_k \qquad (41)$$

where, $\mathbf{z}_k$ is a 1x1 vector measurement at step k, $\mathbf{h}_k$ is a $1 \times 2n$ vector giving the ideal (noiseless) connection between the measurement and the state vector, $\boldsymbol{\nu}_k$ is 1x1 measurement noise vector assumed to be a white sequence with known covariance structure and uncorrelated with the $\mathbf{w}_k$ sequence.

$$E[\mathbf{w}_k \mathbf{w}_i^T] = \begin{cases} Q_k, & i = k \\ 0, & i \neq k \end{cases} \qquad (42)$$

The noise is usually described by its variance, R_k, where

$$E[\boldsymbol{\nu}_k \boldsymbol{\nu}_i^T] = \begin{cases} R_k, & i = k \\ 0, & i \neq k \end{cases} \qquad (43)$$

In general, the error covariance matrix, (P_k^-), associated with an a priori estimate, $\widehat{\mathbf{x}}_k^-$, is defined to be:

$$P_k^- = \mathbb{E}[\mathbf{e}_k^- \mathbf{e}_k^{-T}] = \mathbb{E}[(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)^T] \qquad (44)$$

Having an a priori estimate, $\widehat{\mathbf{x}}_k^-$, and the associated error covariance matrix, P_k^-, we now wish to optimally improve the estimate using the measurement $\mathbf{z}_k$. This is achieved by a linear blending of the noisy measurement and the prior estimate according to the following equation:

$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^- + \mathbf{k}_k(\mathbf{z}_k - \mathbf{h}_k \widehat{\mathbf{x}}_k^-) \qquad (45)$$

where, $\widehat{\mathbf{x}}_k$ is the updated estimate, $\mathbf{k}_k$ is the blending factor.

The idea is to find the particular blending factor that yields an optimal updated estimate. This is achieved by forming first the expression for the error covariance matrix associated with the updated estimate as:

$$P_k = \mathbb{E}[\mathbf{e}_k \mathbf{e}_k^T] = \mathbb{E}[(\mathbf{x}_k - \widehat{\mathbf{x}}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_k)^T]. \qquad (46)$$

Now, we wish to find the particular $\mathbf{k}_k$ that minimizes the diagonal elements of the matrix P_k, because these elements represent the estimation error variances of the state vector components. This particular blending factor is called the Kalman gain and is found to be:

$$\mathbf{k}_k = P_k^- \mathbf{h}_k^T (\mathbf{h}_k P_k^- \mathbf{h}_k^T + R_k)^{-1}. \qquad (47)$$

The covariance matrix associated with the optimal estimate may now be computed as shown below:

$$P_k = (I - \mathbf{k}_k \mathbf{h}_k) P_k^-. \qquad (48)$$

Now, there is a means of assimilating the measurement at $\mathbf{t}_k$, by the use of the optimal $\mathbf{k}_k$, $\widehat{\mathbf{x}}_k^-$, and P_k^-. At the next step, we need $\widehat{\mathbf{x}}_{k+1}^-$ and P_{k+1}^- to make an optimal use of $\mathbf{z}_{k+1}$. First, the updated estimate $\widehat{\mathbf{x}}_k$ is projected ahead via the state transition matrix (ϕ_k) to obtain the a priori estimate $\widehat{\mathbf{x}}_{k+1}^-$. Thus,

$$\widehat{\mathbf{x}}_{k+1}^- = \phi_k \widehat{\mathbf{x}}_k. \qquad (49)$$

The error covariance matrix associated with $\widehat{\mathbf{x}}_{k+1}^-$ is then obtained by forming the expression for the a priori error

$$\mathbf{e}_{k+1}^- = \mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1}^- = \phi_k \mathbf{e}_k + \mathbf{w}_k. \qquad (50)$$

Thus

$$P_{k+1}^- = \mathbb{E}[\mathbf{e}_{k+1}^- \mathbf{e}_{k+1}^{-T}] = \phi_k P_k \phi_k^T + Q_k. \qquad (51)$$

It should be noted that the Kalman gain, in the usual linear recursive Kalman filter, is independent of measurements. The Kalman gain vector, which is the key parameter, can be computed off-line.

C FORMAL LANGUAGE THEORY

C.1 Elements of formal language

An alphabet V, is a finite set of symbols. For any alphabet in V, the countably infinite set of all sentences over V including λ is the closure of V, denoted by V^*. The positive closure of V is the set $V^+ = V^* - \lambda$.

An *empty string* is denoted by λ such that for any $z \in V$,

$$\lambda z = z\lambda = \lambda z. \qquad (52)$$

The *null string* nullifies when it is used in concatenation

$$\Phi x = x\Phi = \Phi \qquad (53)$$

C.2 Grammar and Language

A basic system studied in formal language theory is one that gives a finite set of rules for generating exactly the set of strings in a specific language. These rules of syntax are generally embodied in a *grammar* defined formally as a four-tuple $G = (N, T, P, S)$ where:

- N is a finite set of *non terminals* or variables
- T is a finite set of *terminals* or constants
- P is a finite set of *production* or rewriting rules, and
- S in N is the *starting* symbol.

It is required that N and T be disjoint sets; that is $N \cap T = \Phi$, the null set. The alphabet V of the grammar is the set $N \cup T$. The set P of productions consists of rewriting rules of the form $\alpha \rightarrow \beta$, where α is in V^*NV^* and β is in V^*, with the physical interpretation that string α may be written as, or replaced by, β. α must contain at least on nonterminal. The nonterminals are denoted by capital letters. Lower case letters are used for the terminals.

The *language generated by G*, denoted by *L(G)*, is the subset of T^* obtained by starting with the nonterminal S and using a finite number of productions. In set notation,

$L(G) = \{x | x \text{ in } T^*, S \stackrel{G}{\rightarrow} x\}$

A context-free grammar is the *Greibach normal form* if each of its productions is of the form $A \rightarrow a\alpha$ for nonterminal A, terminal a, and α in N^*.

REFERENCES

Agrawal, R., Lin, K.-I., Sawhney, H. S., and Shim, K. (1995). Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. In *Proceedings of the International Conference on Very Large Data Bases*, pages 490–501, USA. Morgan Kaufmann Publishers Inc.

Amengual, J. C. and Vidal, E. (1998). Efficient Error-Correcting Viterbi Parsing. *IEEE Transactions Pattern Analysis Machine Intelligence*, 20(10):1109–1116.

Angrisani, L., Daponte, P., and D'Apuzzo, M. (1998). A method based on wavelet networks for the detection and classification of transients. *Proceedings of IEEE Conference on Instrumentation and Measurement Technology Conference*, 2:903–908.

Barham, P. M. and Humphries, D. E. (1969). Derivation of the kalman filtering equations from elementary statistical principles. Technical Report TR69095, RAE.

Bollen, M. H. J., Gu, I. Y. H., Axelberg, P. G. V., and Styvaktakis, E. (2007). Classification of underlying causes of power quality disturbances: deterministic versus statistical methods. *EURASIP Journal of Applied Signal Processing*, 2007(1):172–188.

Brown, R. G. and Hwang, P. Y. C. (1992). *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley, New York, 2nd edition.

Chodorowski, J. and Miclet, L. (1998). Applying grammatical inference in learning a language model for oral dialogue. In Honavar, V. and Slutzki, G., editors, *Proceedings of the International Colloquium on Grammatical Inference*, volume 1433, pages 102–113, USA. Springer.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.

Crossman, J. A., Guo, H., Murphey, Y. L., and Cardillo, J. (2003). Automotive Signal Fault Diagnostics Part I: Signal Fault Analysis, Signal Segmentation, Feature Extraction and Quasi-Optimal Feature Selection. *IEEE Transactions on Vehicular Technology*, 52(4):1063–1075.

Forney, J. G. D. (1973). The Viterbi Algorithm. In *Proceedings IEEE*, volume 61, pages 268–278.

Gaouda, A. M., Kanoun, S. H., Salama, M. M. A., and Chikhani, A. Y. (2000). Wavelet-based intelligent system for monitoring non-stationary disturbances. In *Proceedings of the International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 84–89.

Girgis, A. A., Chang, W. B., and Makram, E. B. (1991). A digital recursive measurement scheme for online tracking of power system harmonics. *IEEE Transactions on Power Delivery*, 6(3):1153.

Girgis, A. A., Fallon, C. M., Rubino, J. C. P., and Catoe, R. C. (1993). Harmonics and transient overvoltages due to capacitor switching. *IEEE Transactions on Industry Applications*, 29(6):1184.

Heydt, G. T. (1998). Electric power quality: A tutorial introduction. *IEEE Computer Applications in Power*, 11(1):15–19.

Hochheiser, H. and Shneiderman, B. (2002). A Dynamic Query Interface for Finding Patterns in Time Series Data. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 522–523.

IEEE (1995). IEEE recommended practice for monitoring electric power quality. *IEEE Std 1159-1995*.

Ji, T. Y., Lu, Z., and Wu, Q. H. (2008). Detection of power disturbances using morphological gradient wavelet. *Signal Processing*, 88(2):255–267.

Kalman, R. E. (1960). A new Approach to Linear Filtering and Prediction Problems. *Transactions ASME, Journal of basic engineering*, (82):34–45.

Kezunovic, M., Abur, A., and Huang, G. (1998). MERIT 2000-a new concept in power engineering education. In *Proceedings of the International Conference on Energy Management and Power Delivery*, volume 1, pages 54–59.

Kim, C., Shin, J. H., Yoo, M.-H., and Lee, G. W. (1999). A study on the characterization of the incipient failure behavior of insulators in power distribution line. *IEEE Transactions on Power Delivery*, 14(2):519–524.

Kim, C. J. (2000). Identification of symptom parameters for failure anticipation bytimed-event trend analysis. *IEEE Power Engineering Review*, 20(9):48–49.

Kim, C. J., Lee, S.-J., and Kang, S.-H. (2004). Evaluation of feeder monitoring parameters for incipient fault detection using Laplace trend statistic. *IEEE Transactions on Industry Applications*, 40(6):1718–1724.

Lu, Z., Smith, J. S., Wu, Q. H., and Fitch, J. (2006). Identification of power disturbances using the morphological transform. *Transactions of the Institute of Measurement and Control*, 28(5):441–455.

Luebke, K. and Weihs, C. (2005). Improving feature extraction by replacing the fisher criterion by an upper error bound. *Pattern Recognition*, 38(11):2220–2223.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition.

Manivannan, K. M. and Pankajakshan, P. (2003). Masking of Arcing Related Transients on a Radial Distribution Feeder Due to Capacitor Banks. Technical report, Texas A&M University, College Station, USA.

Miclet, L. and Quinqueton, J. (1988). Learning from examples in sequences and grammatical inference. In Ferrate, G., Pavlidis, T., Sanfeliu, A., and Bunke, H., editors, *Syntactic and Structural Pattern Recognition (Proceedings of the NATO Advanced Research Workshop)*, volume 45 of *NATO ASI Series F*, pages 153–171. Springer, Barcelona.

Morrill, J. P. (1998). Distributed recognition of patterns in time series data. *Communications of the ACM*, 41(5):45–51.

Pankajakshan, P. and Kumar, V. (2007). Detail-preserving image information restoration guided by svm based noise mapping. *Digital Signal Processing*, 17(3):561–577.

Rulot, H., Prieto, N., and Vidal, E. (1989). Learning accurate finite-state structural models of words through the ECGI algorithm. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 643–646.

Rulot, H. and Vidal, E. (1987). Modeling (sub)string-length based constraints through a Grammatical Inference method. In Devijuer and Kittler, editors, *Pattern Recognition Theory and Applications (Proceedings of the NATO Advanced Research Workshop)*, volume 30 of *NATO ASI Series F*, pages 451–459. Springer.

Rulot, H. and Vidal, E. (1988). An efficient algorithm for the inference of circuit-free automata. In Ferrate, G., Pavlidis, T., Sanfeliu, A., and Bunke, H., editors, *Syntactic and Structural Pattern Recognition (Proceedings of the NATO Advanced Research Workshop)*, volume 45 of *ASI Series F*, pages 173–184. Springer, New York, NY, USA.

Sabin, D., Brooks, D., and Sundaram, A. (1999). Indices for assessing harmonic distortion from power qualitymeasurements: definitions and benchmark data. *IEEE Transactions on Power Delivery*, 14(2):489–496.

Santoso, S., Powers, E. J., and Grady, W. M. (1994). Electric power quality disturbance detection using wavelet transform analysis. In *Proceedings of the International Symposium on Time-Frequency and Time-Scale Analysis*, pages 166–169.

Vidal, E., Rulot, H., Valiente, J. M., and Andreu, G. (1993). Application of the error-correcting grammatical inference algorithm (ECGI) to planar shape recognition. In *Proceedings of IEE Colloquium on Grammatical Inference: Theory, Applications and Alternatives*.

Wagner, T. (1971). Convergence of the nearest neighbor rule. *IEEE Transactions on Information Theory*, 17(5):566–571.