Electric Power Systems Research 79 (2009) 144–151

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Classification of sags gathered in distribution substations based on multiway principal component analysis

Abbas Khosravi, Joaquim Meléndez*, Joan Colomer

eXit Group, Department of Electronics, Informatics & Automatics, Institute of Informatics and Applications, University of Girona, Girona, Spain

ARTICLE INFO

Article history:
Received 21 March 2007
Received in revised form 28 March 2008
Accepted 24 May 2008
Available online 7 July 2008

Keywords:
MPCA
Voltage sag
Classification

ABSTRACT

Voltage sags, whether they occur in transmission or distribution systems, may severely damage the loads connected to the power system. As these problems could cost a great deal financially, electric utilities are very interested in finding the origins of sags, that is, whether they have been originated in the transmission network (high voltage (HV)) or in the distribution system (medium voltage (MV)). In addition to the needs of utilities and regulators, many researchers have been prompted to develop reliable methods to properly classify sags. Several of these methods, based on classifying meaningful features extracted from data and waveforms, have been proposed in the literature. Unlike those methods, though, we propose a systematic transformation of data, based on multiway principal component analysis (MPCA), to develop a new voltage sag classification procedure. Sampled voltage and current waveforms of previously registered sags are used together with the MPCA technique to obtain a lower dimensional model. This model is then used to project new sags and classify them according to their origin in the power system. Different classification criteria and parameters are examined to maximize the classification rates of not yet seen sags. Applying the proposed method to real sags recorded in substations demonstrates its applicability and power.

1. Introduction

From generators to customers, voltage waveforms can undergo alterations that affect power quality. The most important phenomena that affect voltage quality are short interruptions, voltage dips (or sags), flickers, supply voltage variations and harmonic distortion. High levels of these phenomena result from both network–customer interactions and the inherent reliability and robustness of the network. These disturbances can originate in either the network or the customers' installation affecting the quality of power. Customers provoke disturbances by starting large motors, using frequency converters and rectifiers or operating specific equipment such as melting furnaces, the effects of which depend basically on the short circuit power at the connection point. On the other hand, disturbances originating in the network are caused, for example, by capacitor bank or line switching, faults (short circuits and protection firing), and other external causes such as lightning or short circuits provoked by vegetation and animals. Thus, network design (relaying and grounding) and operation (protection strategies) are very important for voltage quality. Propagation of these disturbances through the network can negatively affect both customers' equipment (they can be responsible for equipment failure and malfunction or production line breakdowns) and network components (i.e. isolators, protections), and considerably reduces the expected lifetimes and increases the maintenance costs. Due to the importance of power quality, regulators have imposed quality norms related to medium and low supply voltage. The majority of these norms are based on the EN 50160 definitions, but some countries have introduced more restrictive variations for supply voltage parameters. One consequence of this new framework has been the proliferation of power quality monitors installed at the customer end and in substations (the number and location vary considerably from one country to another and depends on the design and purposes of the monitoring campaign). The main motivation for utilities to install these instruments is to gain better knowledge of how networks are performing (according to these standards) and provide regulators with a realistic framework for applying new power quality standards over the next years. At the same time customers could be consulted about installing power quality monitors at critical points of the network to analyse the origin of disturbances affecting sensitive customers and propose mitigation actions.

Monitoring the huge amount of data collected by power quality monitors requires more efficient paradigms. These paradigms should be capable not only of recording disturbances at fixed points of the network, but also of analysing the information generated for equipment and providing operators with enough significant information to avoid manually consulting the database and to know

* Corresponding author. Tel.: +34 972 418883; fax: +34 972 418976.
E-mail address: quimmel@eia.udg.edu (J. Meléndez).

0378-7796/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.epsr.2008.05.014

how the network is performing. Thus, a methodological exploitation of historical data, collected at one point of the network, can be used to obtain behavioural models of the system. The availability of records from several synchronized power quality monitors installed at strategic points of the network can facilitate disturbance propagation analysis, which is necessary to improve the monitoring ability of the whole network without increasing the number of monitors. Incident records from the network control centre are used together with events recorded by power quality monitors to explore cause-effect relationships and improve the diagnosticability of the network. These are the new goals proposed by utilities to develop efficient monitoring systems. The strategies to achieve them must focus on data mining principles and use machine learning paradigms.

Several authors have recently addressed this issue from different points of view using different techniques. In [1], for instance, the authors are interested in voltage sag origins and categorize them in three classes using certain features fed to a fuzzy system. A new method to locate the source of voltage sags in a power distribution system using the polarity of the real current component relative to the monitoring point was introduced in [2]. Other works, such as [3], emphasize the use of preprocessing mechanisms (adaptive wavelet networks, AWN) to extract useful features and automatically classify disturbances into four classes called harmonics, voltage sags, voltage swells and voltage interruptions. In an attempt to categorize transient disturbance waveforms in a power system, the authors in [4] present a hybrid scheme using a Fourier Linear combiner, which exploits expert knowledge in a fuzzy expert system. A comprehensive, in-depth review of voltage sags as well as their stochastic assessment is given in [5].

In this paper we focus on using statistical methods to model sags recorded in distribution substations and locate them in either the distribution (medium voltage, MV) or transmission (high voltage, HV) system. The work has been done using a database of sags characterized by voltage and current waveforms recorded in several distribution substations associated with incidents that occurred in the network and were reported by the network control centre (for details see [21]). From this database HV and MV models, with sags originated in the HV and MV network, respectively, have been created and tested using data mining principles. An extension of principal component analysis (PCA) known as multiway PCA (MPCA) has been used to cope with the dimensionality problem resulting from working with waveform records. Previous work has addressed the problematic classification of transmission and distribution voltage sags based on extracting significant features from waveforms. For example, a phasorial analysis and an unsupervised method were compared in [6]. Extracted temporal descriptors, such as sag duration and depth, and their fall and recovery slopes were used with a learning algorithm for multivariate data analysis (LAMDA) in [6–8]. Other attributes used for the same purpose include phase analysis to obtain the initial phase angle shift, the phase angle difference between current and voltage or the power factor angle [2,9].

Although feature extraction is a necessary preprocessing stage in data mining, selecting and combining them to feed a learning algorithm is not always straight forward. An increasing number of attributes (features) do not guarantee better results, but rather the contrary. For this reason the proposal presented in this paper avoids feature extraction in favour of compressing information contained in whole waveforms based on covariance analysis following PCA principles.

The remaining part of this paper is organized as follows. The formal definition of sags and an overview of the problem are presented in Section 2. The MPCA method is clearly presented in Section 3. As the main body of the paper, Section 4 details the proposed method. Numerical results using real data gathered from distribution substations in Spain and interpretation of these data are given in Section 5. Finally, Section 6 ends the paper with a conclusion and some rules of thumb for future work.

2. Problem overview, motivation and objectives

Among the possible disturbances propagated through the network, voltage sags are the most significant due to their severity and number of occurrences per year. Voltage sags that last for only a 100 ms may force an industry to be inoperable for several hours, which implies large costs for companies due to a subsequent loss in production. The formal definition of a voltage sag is a momentary decrease (10–90%) in the root mean square (RMS) voltage magnitude for a duration of 10 ms to 1 min [8]. Several documents related to voltage sags characterize them using two parameters: magnitude and duration [11–15]. While the sag magnitude is defined as the minimum RMS value obtained during the event, its duration is the time interval between the time when the RMS voltage crosses the voltage sag threshold (0.9 p.u.) and the instant when it returns over this value [7]. Sags could affect one, two or three phases. From this perspective sags can be grouped in three classes, phase to ground, phase to phase and three phase, and can originate in either the transmission (HV) or distribution network (MV). Customers can also generate sags, however this study has not contemplated this situation, focusing instead on data gathered in distribution substations.

The main objective of this paper is to develop a procedure for determining whether a sag has taken place in the transmission or distribution network; that is, whether it is HV or MV. This classification is motivated by

- *Rapid fault location in distribution systems*. Faults in the distribution system are generally observed in the substation as sags, whose measurements during the fault are used to estimate their location based on the concept of apparent impedance. Thus, after detecting the presence of a sag in the substation and before running the fault location algorithm, it is necessary to determine if it originates in the distribution network (MV) or comes form the transmission system (HV).
- *Network monitoring*. Some regulators ask utilities for global monitoring of the network. Since it is impossible to install power monitors in all customer locations and/or substations, utilities need strategies to install a minimum number of measurement points while assuring maximum network coverage. This means that not only are the disturbance records (waveform, time stamp, etc.) important, but also knowing the way they are propagated (origin and propagation direction) is essential to knowing how the disturbances can affect other monitored substations.
- *Responsibility for bad power quality*. Distribution companies are responsible for the quality of power supplied to customers, but not all the causes of this poor quality are found in the distribution network. Disturbances that originate in generation and transmission systems are also propagated through the distribution network. Since distribution companies are responsible for the quality received by the customer, they will obviously be interested in distinguishing between disturbances generated in the distribution system and those only propagated by it.

The number and severity of voltage dips, swells and rapid voltage changes depend on several factors, such as [20]: the extension and meshing of the network (HV, MV and LV), the presence of cables or overhead lines, neutral grounding systems (isolated, solidly or