Electrical Power and Energy Systems 71 (2015) 373–382

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

An efficient PD data mining method for power transformer defect models using SOM technique

V.P. Darabad [a], M. Vakilian [a,*], T.R. Blackburn [b], B.T. Phung [b]

[a] Electrical Engineering Department, and Center of Excellence in Power System Management and Control, Sharif University of Technology, Iran
[b] School of Electrical Engineering & Tele., University of New South Wales, Australia

ARTICLE INFO

Article history:
Received 13 December 2014
Received in revised form 28 February 2015
Accepted 17 March 2015
Available online 9 April 2015

Keywords:
Data mining
Power transformer
Defect model
Partial discharge
SOM

ABSTRACT

Suggestion and application of a set of new features for on-line Partial Discharge (PD) monitoring, where there is no information about the type of PD is a challenging task for condition assessment of power equipments, such as a power transformer. This is looked for in this paper. So far, in past various techniques have been employed to develop a comprehensive PD monitoring system, however limited success has been achieved. One of the challenging issues in this field is the discovering of proper features capable of differentiating the involvement of possible types of PD sources. In order to examine the efficiency of the method established in this paper, which is based on application of a set of new feature spaces, texture feature analysis, followed by application of principal component analysis (PCA) and self-organizing map (SOM) is used to analyze and interpret the time-domain-captured PD data. The results of this work demonstrate the capabilities of the aforementioned features space to be used as a supplementary knowledge-base to help experts making their decisions confidently.

Introduction

A large population of power transformers along with other power system grid infrastructures have been in service for decades and considered to be in their final ageing stage. On the other hand, due to economy and business growth in our era, the electricity demand is growing rapidly. Therefore, transformers became the most critical devices in power system due to their long repair or replacement time; so different preventive measures (i.e., traditional maintenance schemes such as corrective and time-scheduled maintenance) are required to ensure that a power transformer is reliable. Today, the traditional maintenance schemes cannot satisfy the requirements of modern power systems (i.e., high reliability and sustainability of electric energy supply). One of the best measures to overcome these concerns is condition monitoring of transformers. Since, the insulation is the most critical part of transformers and other high voltage equipment in their operation; this work employs the insulation monitoring schemes.

Considering the limitation of off-line condition monitoring methods, nowadays power utilities tend to employ the on-line condition monitoring schemes to avoid the cost of traditional maintenance procedures and unnecessary service outages [1,2]. There are three major on-line conditions monitoring technique based on application of DGA [3], DDF [4] and PD [5] measurements. Among these, DGA can only be applied to the transformer oil monitoring. Additionally, if the DGA and DDF methods are employed, the related accurate measurement devices are relatively expensive. PD measurement does not have these limitations and there is a good economic balance between a measuring equipment cost and its accuracy [6].

PD monitoring is a non-destructive test and can be performed continuously by installing the PD measuring equipment in field but it generates huge amount of PD data. Advancement of technology made digital recorders capable of recording PD data in wide time-windows in range of milliseconds and with high sampling rates, i.e. hundreds of mega samples per second. Considering this fact, any decision making system faces huge amount of data in range of gigabytes. To manage this huge amount of data and extract the underlying useful information from such a huge database, the use of traditional data handling techniques such as hierarchical minimum distance classifiers or dendrograms is not

Abbreviations: BMU, best match unit; Ch, channel; DDF, dielectric dissipation factor; DGA, dissolved gas analysis; DWT, discrete wavelet transform; FRA, frequency response analysis; GLCM, gray level covariance matrix; GLDV, gray level difference vector; Gs/s, giga sample per second; HFCT, high frequency current transformer; HV, high voltage; IR, insulation resistance; Lab, laboratory; Mpts, mega points; Ms/s, mega sample per second; N_g, gray level value; p.f., power frequency; PC, principal component; PCA, principal component analysis; PD, partial discharge; PDC, polarization and depolarization current; PRPD, phase resolved partial discharge; RVM, recovery voltage measurement; SOM, self-organizing map; SVM, Support Vector Machine; SNE, stochastic neighbor embedding.

* Corresponding author. Tel.: +98 9121482357.
E-mail address: vakilian@sharif.ir (M. Vakilian).

http://dx.doi.org/10.1016/j.ijepes.2015.03.016
0142-0615/

helpful and promising. However, data mining techniques can be employed to describe similar or different patterns or the relationship between some data within the database [7]. An important concern in application of the PD data mining techniques is to find the features comprehensively capable of differentiating among the involved PD sources. The research works are carried out (in this field) to evaluate the fidelity of feature extraction methods which can be applied on the recorded PD data spaces. Any success in this field leads to the development of a more confident decision system.

To address this concern, the texture analysis technique is applied on the input data space, and a well-known clustering algorithm named "Self-Organizing Map" (rather than using classification methods) is employed to verify if the similar data are classified in the same group/cluster.

The PD condition monitoring techniques will be employed to access the condition of power equipment installed in a power system plant (assuming no knowledge exists about the kind of PD involved).

This approach is implemented into the simulation and categorization of the laboratory built-in PD defects to be verified against them. This will provide a basis for application of the method and characterization of PD types in power transformers.

This paper is organized in eight sections as follows. Section 'Related works' is assigned to literature review. Section 'PD Measurement setup' describes the experimental setups which are used in this work. In section 'Data mining process', the data mining process is described and the related techniques employed for this purpose are discussed. Section 'PD classification model' explains the PD classification method employed in this work. In section '**PD data analysis**', PD data analysis method is applied on the captured data; in section 'Analysis of results' the results of work are analyzed, while section eight is devoted to summary and conclusion.

Related works

In some of the past researches [8], the discharge characteristics are studied to identify the sources of discharges. Different discharge quantities such as discharge numbers, maximum discharge magnitude, inception voltage, etc. are defined in terms of the (n, q, ϕ) domain. Artificial PD models are employed to simulate the PD patterns. Using some of their related features, the obtained databases are analyzed. Some artificial insulation defects are developed [9,10]. Three distributions, for positive and negative half of the voltage cycle (pulse count $H_n \pm (\phi)$, mean pulse height $H_{qn} \pm (\phi)$ and maximum pulse height $H_{q\max} \pm (\phi)$), are defined on the output of PD measurement routine. Based on this knowledge, some industrial cases are investigated and results showed that the proposed method can help users identifying the types of insulation defects in HV equipment. The proposed approach followed in [11] to make a PD database for turbogenerators. It is shown that the proposed PD data base is useful to make a good distinction between turbogenerators which are in good condition and those which showing PD activity originated from their degraded insulation.

In [12], a proposed RF method used to measure the discharge process radiated energy, while IEC 60270 technique measures apparent charge simultaneously on some common PD topologies built in HV lab. It was concluded that the correlation between apparent charge and the energy of the RF signals demonstrates the specific patterns related to each type of PD source.

In [13], a phase resolved partial discharge (PRPD) pattern is recorded using partial discharge sensors used to capture the PD data related to an oil-filled power transformer. Similar to [12], discharge distributions are defined and statistical features are extracted from them. Based on these features a knowledge-rule-base is developed to diagnose power transformer PD defects.

In [14], statistical features are extracted from PRPD patterns of three basic types of PD, corona, surface discharge and internal discharge; a combination of genetic algorithm and decision tree is employed as a descriptive data mining tool to mine rules which are useful to differentiate PD types.

In [15], to find the location of single and double PD sources, optical sensors were used. And, five different energy features are extracted from time-domain recorded PD signals. By application of PCA; best features were selected and fed into SVM-classifier.

In [16], to recognize the insulating defects in HV XLPE cables, a feature extraction method is employed through application of gray intensity image of PD.

Authors in [17] applied a feature extraction method on the recorded PD pulses through S-transform [18]. The feature size-reduction was performed by using none-negative matrix factorization. The feature selection was followed by NSGA-II algorithm [19] and classification was done using FCM algorithm.

In [20], to localize the PD source in power transformer, features are extracted from the time-domain PD recorded data (of different artificial PDs types applied into power transformer windings) using karhunen–loeve, discrete cosine and sine; hadamard, and haar transform.

In [21], to classify the PD sources, feature extraction method including SNE, DWT and static moments; are applied on PRPD and Time-domain PD data. Additionally, different classification algorithms like FSVM are used to classify single and multi event PD sources.

In [22], feature extraction is implemented, through application of wavelet packet decomposition technique on UHF PD signals detected by a Hilbert fractal antenna. SVM and BPNN were used to classify the six PD sources were involved.

In [23], extensional neural network was used as a PD classifier. Features which are fed into this classifier were extracted from a set of raw three-dimensional PD data.

In [24] orthogonal least square center selection technique was employed to recognize multiple source PD patterns. In this work radial basis probabilistic neural networks were compared with the standard and the heteroscedastic probabilistic neural networks.

As it is mentioned in introduction, finding proper features is the key factor in the success of the PD discrimination task. Having a few decision systems, or on the other hand, involvement of various feature spaces, results in better and more reliable resolutions under tough situations.

A review on literature emerges the fact that in (n, q, ϕ) method, usually a "narrow band coupler" is used to decrease the noise effect which in contrast puts a limit on frequency content of the recorded data (since all the information obtained from this method is just related to a few discharge quantities), and there would be no information about the PD pulse waveform. On the other hand, the single PD signal shape is usually employed to locate PD events, not to categorize PD types.

In order to maintain the benefits of the (n, q, ϕ) method in PD signal waveform analysis, measurements are performed in the ultra-wide frequency to preserve the information of each single PD activity [25], and the whole power frequency cycle is considered as an input vector to preserve the information obtained in the (n, q, ϕ) domain. In Authors previous work [28], the sampling rate of the recording device was set to 1 Gs/s which produced extremely large input vectors and consequently slowing down the feature extraction and the classification process. In this work, the sampling rate has been reduced to 100 Ms/s to improve the speed of data mining process.