Fuzzy Data Mining as a Tool to Infer Pollution Severity of Power Transmission Line System

K. Karpagavani\(^1\)* and A. Kumaravel\(^2\)

\(^1\)Department of Sciences, Sona College of Technology, Salem-636005, Tamilnadu, India
\(^2\)Department of Mechanical Engineering, K.S. Rangasamy College of Technology, Tiruchengode-637215, Tamilnadu, India

Abstract

Flashover of power transmission line due to polluted insulators leads to failure of equipments, catastrophic fires and power outages. This paper deals with the development of pollution severity diagnostic system of power transmission line insulators based on the fuzzy data mining approach. In this work, laboratory based experiments were carried out on power transmission line porcelain insulators under AC voltages at different pollution conditions and corresponding leakage current patterns were monitored. Variation of important features of leakage current such as leakage current magnitude, total harmonic distortion and third harmonic content at different pollution levels were analyzed. Fuzzy C-means algorithm is used to cluster the extracted features of the leakage current data. Test results clearly shows that the fuzzy clustering technique can effectively realize the surface pollution condition of the power transmission line insulators and can be useful to carry out preventive maintenance work.

Key Words: Insulator, Power Systems, Leakage Current, FFT, THD, Fuzzy C-Means

1. Introduction

Flashover of high voltage power transmission line due to contaminated insulators is a major trouble faced by electrical utilities, which results in power outages, waste of time and money, cause frustration to customers, equipment damage and potentially, catastrophic fires. Therefore, utilities spend significant amount of money on preventive maintenance, which includes insulator washing and cleaning, but it is an expensive operation and difficult to automate [1,2]. Various approaches are used to quantify the pollution severity of power transmission line systems in order to carry out preventive maintenance work [3,4]. Leakage current (LC) in the contaminated insulator plays a major role in the flashover of power transmission line. Therefore, measurement and analysis of LC pattern will provide useful information for the development of efficient diagnostic system for power transmission line.

Nowadays, development of condition monitoring systems using the data acquisition systems and control is a basic tendency in automation of power system equipments and control [5]. Intensive growth of the information systems which are based on the internet and web-technologies creates the possibility to access to these condition monitoring systems from anywhere in the world. Development of the web-based diagnostic system using the data acquisition systems for several practical applications is a vital and hot research issue. Therefore concept of remote management of high voltage power transmission line insulators over the internet from anywhere in the world can be a reality with existing technologies. In addition, the use of the internet technology reduces exploitation costs of existing communication channels. However, collection of LC signal using data acquisition systems over a period of time results in a
huge amount of data, which makes it difficult to implement the web based technologies for the condition monitoring system [6].

In this regard, extraction of important features from the LC data and usage of effective data mining techniques to predict flashover phenomena are expected to play important roles in the design of web based diagnostic system for power transmission lines. Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to extract meaningful patterns and useful information [7]. Clustering is the method of grouping objects into meaningful subclasses so that the members from the same cluster are quite similar [8]. As the leakage current of power transmission system follows different patterns during the various stages involved in the development of flashover process, data mining using fuzzy clustering methods can be useful for accurate prediction of contamination severity of power transmission system.

Considering the above facts, the aim of the present work is to develop a structured method of analyzing the LC data using fuzzy data mining approach for the development of efficient diagnostic system for power transmission lines. A due emphasis is given to monitor the important parameters such as leakage current magnitude, third harmonic content, total harmonic distortion etc., in this paper.

2. Experimental Setup

Figure 1 shows the overall structure of the proposed web based diagnostic system for power transmission lines. Leakage current is the important parameter measured from the high voltage transmission line tower insulator. It is then processed in the data acquisition system and stored in the server for further processing. The client will have the access to the data in the server through TCP/IP connection.

Figure 2 shows the schematic diagram of the laboratory experimental setup used for implementing the web based diagnostic system. Tests were conducted as per IEC 60507 clean fog test procedure [9]. The single disc porcelain insulator was suspended vertically inside the fog chamber. A 100 kV high voltage test transformer with control panel was used to supply the required test voltage. The test voltage was maintained at 11 kV rms, 50 Hz. To reproduce saline pollution typical of coastal areas, a contamination layer consisting of NaCl and 40 g of kaolin mixed with 1 litre of deionized water was applied to the surface of insulator [10]. Four ultrasonic nebulizers were used to maintain the required relative humidity level inside the fog chamber.

3. Leakage Current Measurement

The leakage current was measured using a Pearson High Frequency Current Monitor Q-1903 in the ground lead. It has a sensitivity of 1 V/A, lower cut off frequency of 15 Hz and higher cut off frequency of 500 kHz. A high sampling rate data acquisition system (National Instruments, USB 6251, 1.25 MSa/sec) was used in the present study to access analog and digital signal. This system is capable of measuring 16 analog input signals, 12 or 16 bits. A LabVIEW software system developed in the server for this data acquisition system provides the user with the complete LC waveforms, which are therefore available for further signal processing.

4. Selection of Features

Extraction of salient features of the LC data, which

![Figure 1. Overall structure of the system.](image1)

![Figure 2. Schematic diagram of experimental setup.](image2)
in turn actively drives diagnostic knowledge out of the raw data, plays a major role in the novel condition monitoring technologies. In order to develop an efficient diagnostic system, it is necessary to perform both time and frequency domain analysis of LC signals. In the case of time domain analysis, magnitude of LC plays a major role in determining the surface condition of the insulator. In the case of frequency domain analysis, important features of LC are Total Harmonic Distortion (THD) and variation of third harmonic content with respect to time \[11,12\]. Fast Fourier Transform (FFT) technique has been found to be efficient to extract features from the leakage current data. In this work, feature extraction of LC signal is carried out using Fast Fourier Transform technique, which is an effective tool to understand frequency characteristics of LC signals.

5. Data Mining Using Fuzzy C-Means Clustering

In recent times, for accurate analysis and classification of signals with complex characteristics, fuzzy c-means clustering technique has been proposed as an effective tool \[13,14\]. The fuzzy c-means (FCM) algorithm has admirable popularity in a great number of fields. Fuzzy c-means is a method of clustering which allows one elements of the data set to belong to two or more clusters. In this method, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. Therefore, in the present work, fuzzy c-means clustering technique has been adopted for diagnosing the surface contamination condition of the insulators.

Fuzzy c-means algorithm is based on the concept of fuzzy C partition. Let \(X_i \in \mathbb{R}^p\), \(i = 1, ..., N\), denote the data elements represented as \(n\) real-valued columns vectors of dimension \(p\). Let \(C_j \in \mathbb{R}^p\), \(j = 1, ..., C\), represent the center of cluster, with \(2 \leq C < N\). Let \(U \in \mathbb{R}^{C \times N}\) denote the partition matrix comprised of fuzzy memberships. The elements of \(U\) satisfy the following constraints

\[
\begin{align*}
0 \leq u_{ij} & \leq 1 \\
\sum_{j=1}^{C} u_{ij} & = 1
\end{align*}
\]

The fuzzy c-means clustering is based on the following optimization function, under the constraints in (1)

\[
\min_{u, c} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij} \left\| X_i - C_j \right\|^2, 1 \leq m \leq \infty \right\} \quad (2)
\]

where \(m\) is any real number greater than 1, it controls the amount of “fuzziness”, \(u_{ij}\) is the degree of membership of \(X_i\) in the cluster \(j\), and \(\| \cdot \|\) is any norm expressing the similarity between any measured data and the center. The cluster centers \(C\) can be measured by

\[
C_j = \frac{\sum_{i=1}^{N} u_{ij}^m X_i}{\sum_{i=1}^{N} u_{ij}^m} \quad (3)
\]

Fuzzy partitioning is carried out through an iterative optimization of the objective function, with the update of membership \(u_{ij}\) and the cluster centers \(C_j\) by

\[
u_{ij} = \left[ \sum_{k=1}^{C} \left( \frac{\left\| X_i - C_k \right\|}{\left\| X_j - C_j \right\|} \right)^{2(m-1)} \right]^{-\frac{1}{m-1}} \quad (4)
\]

The following algorithm finds a solution that converges to a local minimum of (2) for the FCM method,

Step 1: Initialize \(U = \{ u_{ij} \}\) matrix, \(U^{(0)}\)

Step 2: At \(k\)-th iteration: calculate the center vectors \(C^{(k)} = \{ C_j \}\) with \(U^{(k)}\), using (3).

Step 3: Update \(U^{(k)}\) to be \(U^{(k+1)}\) using (4).

Step 4: Stop rule: If \(\max_{ij} \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| < \varepsilon\) where \(\varepsilon\) is a small number between 0 and 1, then stop. Otherwise, \(k = k + 1\), and return to Step 2.

6. Structure of the Diagnostic System

Figure 3 shows the schematic diagram of the diagnostic system developed for the pollution severity monitoring of the transmission line insulators. LC signal is processed in both time and frequency domain and the important features such as LC\(_{\text{peak}}\) and THD are extracted from the LC data. Extracted features are given as an input to the fuzzy c-means clustering algorithm and then output of this module is given to the pollution severity meter, which indicates the level of pollution of the insulator to the substation operator.
Initially, time domain analysis of the trend followed by the peak value of the LC is carried out over a period of time. In this case, test voltage is applied continuously to the insulator specimen at a constant pollution level in the fog chamber and LC peak is captured. Figure 4 shows the trend followed by the peak value of the LC at medium and heavy pollution condition of porcelain insulator. It is observed from the experimental studies that peak value of LC is highly intermittent in nature and also it is not steadily increasing over a period of time.

It is highly fluctuating and not constant over a time period for a particular pollution level. Peak value of LC varies randomly from 2 mA to 30 mA. It is very difficult to arrive at any decision on the pollution level of the insulator, just only from the results of time domain analysis of peak value of LC at any time.

In order to arrive at any reasonable estimate it is also necessary to perform frequency domain analysis of individual leakage current signals. In the frequency domain method, pollution severity analysis is performed using the trend followed by the Total Harmonic Distortion (THD) of the leakage current signal. Figure 5(a–e) shows the typical LC waveform patterns obtained during experimental studies. For a given insulator, LC waveform evolution depends essentially on the changes occurring at the surface pollution layer and surface wetness of the insulator. From the results, it is clear that there is a significant increase in the magnitude of the leakage current with increase in pollution level. Also, during heavily
polluted conditions, several short duration discharges were observed. During very heavily polluted conditions, several long arcs were observed which will lead to flashover of the insulator. These LC signals were used for frequency domain analysis using FFT.

In this work, Fast Fourier Transform (FFT) based spectral analysis of LC is developed in the LabVIEW software to understand the harmonic components of signal at various pollution levels. Analysis of third harmonic contents and THD will be useful to distinguish the different stages of development of flashover. THD of leakage current signal is calculated as follows

\[
THD = \frac{\sqrt{\sum_{n=2}^{n} I_{n}^2}}{I_{rms}} \times 100\%
\]  

(5)

where

\[
I_{rms} = \sqrt{\sum_{n=1}^{n} I_{n}^2}
\]  

(6)

and corresponding FFT plots are shown in Figure 6 (a–e).

The following points were observed from the time and frequency domain analysis,

(i) Under lightly polluted conditions, the magnitude of leakage current is small without any surface discharges and the %THD lies in the range of 35–42% (Figures 6a and b). Third order harmonic components are less in the signal.

(ii) Under medium polluted conditions, presence of short duration discharges (which lasts for half or one cycle) as shown in Figure 5c were noticed and the THD value lies in the range of 45–58%. A significant increase in third harmonic content is also noticed.

(iii) Under heavily polluted conditions, the frequency of occurrence of short duration discharges increased considerably (as shown in Figure 5d) and the THD value also reached above 60% with a considerable increase in third harmonic content.

(iv) Whenever there is a severe discharge formation due to very heavy pollution, then discharges were observed for duration of 5-25 continuous cycles (Figure 5e). Under such circumstances, the discharge pattern almost looks like a sinusoidal waveform, with THD value lying in the range of 10–25%. Higher order harmonic components are completely absent and third order harmonic components play a major role in this case.

Since the LC data is captured continuously and it is also highly intermittent in nature, it is necessary to cluster the captured data over a period of time by using recent data mining techniques. As a data mining function, cluster analysis can be used as a stand-alone tool to gain in-

![Typical FFT plot of leakage current signals](image)

Figure 6. Typical FFT plot of leakage current signals (a) & (b) lightly polluted, (c) medium polluted, (d) heavily polluted, (e) very heavily polluted.
sight into the distribution of data, to observe the characteristics of each cluster, and to focus on a particular set of clusters for further analysis. It may also serve as a pre-processing step for characterization and classification of events. In order to identify the groups of similar objects and to discover distribution of leakage current patterns, fuzzy c-means clustering models were built by MATLAB. Fuzzy clustering process stops when the objective function improvement between two consecutive iterations is less than $\epsilon$ (set $\epsilon = 1e-5$), or when the maximum number of iterations (set as 100) is reached.

The LC data over a period of time during the study periods were used for the fuzzy clustering. Four clusters (pollution conditions) are considered in this study, namely, ‘lightly polluted’, ‘medium polluted’, ‘heavily polluted’ and ‘very heavily polluted’. Figure 7(a) show the LCpeak-THD relationships captured at lightly polluted conditions. In Figure 7(b), LCpeak-THD relationships after fuzzy clustering are shown, where the markers ‘*’, ‘x’, ‘+’ and ‘o’ denote four cluster of data corresponding to ‘lightly polluted’, ‘medium polluted’, ‘heavily polluted’ and ‘very heavily polluted’ respectively and these clusters are also marked as ‘1’, ‘2’, ‘3’, ‘4’ respectively. During lightly polluted conditions, since the peak value of LC is small, it is observed that cluster 1 which corresponds to lightly polluted conditions is having large number of data points when compared with cluster 2, 3 and 4. Figure 8 shows the typical plot of fuzzy clustering process convergence criterion with respect to number of iterations under lightly polluted conditions.

Figure 9 shows the LCpeak-THD relationships captured at medium polluted conditions. When compared with lightly polluted conditions, the number of data points in the cluster 2 is slightly increased. However, there is no significant increase in data points corresponding to cluster 3 and 4. Similar plots of LCpeak-THD relationships captured at heavily polluted and very heavily polluted conditions are shown in Figures 10 and 11 respectively. From these figures, it is clear that number of data points in cluster 3 and cluster 4 increases considerably with respect to increase in pollution. This fuzzy clustered LC data plots clearly indicates the surface pollution condition of the insulators. For practical applications, it can be speculated that cluster density above certain threshold value could warrant corrective actions and which will be useful for substation operator.

The correctness of fuzzy c-means clustering algorithm results should be verified by using appropriate criteria and techniques. There are several methods proposed in the literatures to validate the accuracy of the clusters [15]. Measuring the distance between the clusters is a common approach, which is done by measuring the distance between the closest members or the distant members of the clusters. However, measuring the dis-

![Figure 7](image7.png)

**Figure 7.** Typical LCpeak-THD data of leakage current signals under lightly polluted conditions (a) before clustering, (b) after clustering.

![Figure 8](image8.png)

**Figure 8.** Typical plot of fuzzy clustering process convergence criterion with respect to number of iterations under lightly polluted conditions.
tance between the centers of the clusters (or) centroids aims at finding the best clustering scheme. In this paper, measuring the distance between the centroids method is used to measure the cluster accuracy.

Figure 12 shows the centroids of the clusters obtained at four different polluted conditions as discussed earlier. Distance between the centroids of the clusters is denoted as D12, D13, D14, etc. as shown in Figure 12. At each pollution condition, the distance between the centroids of the clusters i and j were calculated using the equation

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(7)

where x and y are the (x, y) coordinates of the respective centroid point. Figure 13 shows the bar chart of the distance between the centroids calculated at four different pollution conditions.

It is observed that the distance between the centroids of the clusters are closely located at each pollution condition. In order to understand the deviation in the distance between the cluster centroids, standard deviation is calculated using the following equation,

$$\text{std}_\text{dev} = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} [D_{ij}(k) - \mu]^2}$$

(8)

where $\mu$ is the mean of the vector $D_{ij}$ and N is the length of the vector $D_{ij}$. The standard deviation values obtained for each distance between the centroids are also shown above corresponding bar chart in Figure 13. It is observed that standard deviation value varies from 0.68 to 2.7, which is very less and within acceptable limit. It clearly indicates that the fuzzy c-means technique is more reliable for clustering the leakage current data of the power transmission line insulators.

From the above reported results, it is noticed that the fuzzy c-means clustering technique is very much useful for easy identification of surface pollution severity of insulators used for high voltage applications. It is also ob-

---

**Figure 9.** Typical LCpeak-THD data of leakage current signals under medium polluted conditions (a) before clustering, (b) after clustering.

**Figure 10.** Typical LCpeak-THD data of leakage current signals under heavily polluted conditions (a) before clustering, (b) after clustering.
served that the leakage current magnitude and THD relationships are directly related with surface pollution severity. This can be easily understood from the cluster plot of insulator obtained at different pollution conditions. The proposed pollution severity diagnostic system results show that it can effectively realize the pollution severity of outdoor transmission line insulators and this system is easily applicable for real time measurements.

7. Conclusion

In this paper, a diagnostic system for the pollution severity analysis of power transmission line insulators using fuzzy c-means clustering technique was proposed. It was developed in such a way to take into account the important features of leakage current of insulator such as LC peak and THD. Test results in this work clearly indicate that it is simple to implement web based technologies for the development of condition monitoring system of power transmission lines using the FFT feature extraction and fuzzy clustering techniques. Based on the fuzzy cluster results, it is easier for the substation operator to take decisions or initiatives for preventive maintenance work.

Acknowledgement

First Author (K. K.) would like to thank the Sona-PERT R&D Centre of Sona College of Technology, Salem-636005 for providing the lab facilities and support during the experimental studies.

References

[1] Gorur, R. S., Cherney, E. A. and Burnham, J. T., Out-
Fuzzy Data Mining as a Tool to Infer Pollution Severity of Power Transmission Line System


Manuscript Received: Mar. 2, 2012
Accepted: Oct. 29, 2012