Hybrid k-means-PSO Technique for Transformer Insulation Moisture Determination in the Production Stage Based on Frequency Response Analysis

M. Bigdeli* (C.A)

Abstract: Moisture in the transformer insulation can shorten its life. There are many methods for detecting humidity in transformer paper insulation. One of the methods used in the factory to evaluate the drying process of transformer insulation and determine its humidity is the frequency response analysis method. In this paper, the desired experiments are performed on different transformers, and after obtaining the results of frequency response measurements, the required features are extracted from them. Then, using the k-means method, these features are placed in three clusters (dry, wet, and excessively wet). The cost function of the k-means method is optimized using the particle swarm optimization (PSO) algorithm to get a better result. By applying new data from different transformers, the capability of the proposed method in determining the moisture content of the transformer is evaluated. The results obtained from the evaluation of the insulation condition of another group of transformers indicate the high accuracy of the proposed method.

Keywords: Frequency Response Analysis, k-means, Moisture, PSO, Transformer.

1 Introduction

POWER transformers are one of the essential equipment of electric power transmission networks. Failure of any of them will reduce the reliability of the network, power outage, and impose high costs of repair and transportation. For this reason, and due to the competitiveness of the electricity industry, today, the improvement of condition monitoring methods and fault diagnosis is of particular importance. This issue has become more critical over time and has entered a newer field so that even medium and small transformers have been discussed. Improving the quality of factory production is essential to increase reliability and prevent waste of capital. Insulation failure is one of the most important defects that causes the transformer to go out of circuit and cut off the transmission of electrical energy. To prevent damage to the transformer insulation, the drying process of the transformer insulation must be appropriately performed during production. In oil-immersed transformers, there are two main insulators, paper, and oil. There are well-known methods such as dissolved gas analysis (DGA) to monitor the condition of transformer oil [1, 2]. However, the main challenge in condition monitoring of the transformer is assessing the transformer paper’s condition because paper sampling is complicated and almost impossible. The biggest enemies of insulation, especially insulating paper, are moisture, heat, and oxygen, which heat and oxygen produce more moisture. Therefore, our most crucial challenge in maintaining the transformer will be the fight against humidity, which is the first step to determine the amount of water [3]. The transformer’s active part (core, windings, and tap changer) is heated and vacuumed during the production stage. Heating and vacuuming remove moisture from the insulation. Therefore, the dryness of paper insulation and moisture reduction is related mainly to the production stage in the factory. However, in rigid bodies, some moisture remains, which increases the operating time of the
transformer and temperature variations; this moisture is removed and enters the oil. Also, according to the cellulose structure of the insulating paper, due to heat breaking of polymer bonds, hydrogen and oxygen are produced, which combine hydrogen and oxygen to form water. This phenomenon is called insulating aging. Therefore, detecting the moisture content of the transformer insulation during the production stage can prevent irreparable damage.

Much research has been done to improve the drying process of transformer insulation. The effect of the low-frequency heating (LFH) method on the drying quality of transformer paper insulation has been investigated [4]. In [5], using the finite element method, the various techniques of transformer field drying that are currently widely used have been analyzed, and the efficiency of these methods has been studied and compared. Recommendations for drying parameters are also provided. A new method based on the analysis of dielectric response in the frequency domain using the artificial bee colony algorithm to determine the moisture content of the transformer insulation system is presented in [6]. Reference [7] presents a new method based on frequency dielectric spectroscopy (FDS) for controlling the drying process of transformer insulation in the factory. Another new method called Mobile Vapor Phase Drying (MVPD) has been proposed for transformer field drying in [8]. Statistical analysis of the effect of temperature on the drying time of transformer insulation in the drying process is presented in [9].

One of the methods used to evaluate the drying quality of transformers is the frequency response analysis (FRA) method. The FRA method, also known as the transfer function (TF) method, is a comparative method [10]. In this method, the customer or the manufacturer keeps transformer experiments in healthy condition as the result of the reference measurements. During the annual inspection or when a fault occurs in the transformer, the same measurement is performed again with the same terminal connection conditions and environmental conditions. If there is a significant difference between the results of the measurements, it should be evaluated to determine if the condition is abnormal. FRA has the ability to detect various defects of the transformer [10]; however, in this paper, its ability to detect the moisture content of the transformer is investigated. In [11], it is shown that FRA is sensitive to the moisture content of the insulation, and it can be used to assess the condition of the insulation. In [12, 13], it has been shown that by measuring the FRA, the drying condition of the transformer insulation can be evaluated.

Despite the valuable studies that have been done to determine the time required for the transformer to dry and to determine the moisture content of its insulation, it is still a significant and fundamental issue to determine the amount of moisture in the transformer during the production stage. Due to the fact that today the FRA method is used as a standard method in condition monitoring of the transformer, so in this paper, this method is used to determine the moisture content in the transformer during the production stage. It is better to use data mining and machine learning methods to interpret the results to increase the efficiency of FRA in transformer condition monitoring. Although valuable research has been done with the help of data mining and machine learning in diagnosing the defects of transformer windings using FRA [14-17], no significant work has been done in the field of moisture detection using FRA-based data mining methods. In this paper, to overcome this shortcoming, the desired experiments are first performed on a 50 MVA, 132/33 kV oil-immersed transformer, and its TFs are measured in different states of the insulation condition (dry, wet, and excessively wet). Then, based on the measurement results, according to the Chinese standard [18], the normalized correlation coefficient (NCC) numerical index is calculated in three frequency ranges of 0-100 kHz, 100-600 kHz, and 600-1000 kHz. By applying the NCC-based extracted features using the k-means method, the insulation condition of the transformer is clustered into three different clusters: dry, wet, and excessively wet. In order to increase the accuracy of the k-means method in data clustering, the particle swarm optimization (PSO) algorithm is used to optimize its objective function. After determining the centers of the clusters with the help of PSO, to evaluate the performance of the proposed method, the data obtained from the other three transformers, whose insulation moisture content is known, are applied to k-means. Analysis of the results shows that the proposed method can accurately detect the transformer’s insulation moisture content.

2 Theory of Methods Employed

In this paper, the k-means method based on the PSO algorithm is used to detect the moisture content of transformers in the factory. The theory of these methods is given in detail in the literature [19-21]. Therefore, they are briefly discussed in this section.

2.1 k-means

Clustering is a fundamental process in data analysis. In clustering, the goal is to divide a set of data into groups called clusters so that ideally, objects in one group are precisely the same and objects in other groups are different. The k-means method is one of the simplest and most popular clustering algorithms used in data mining, especially unsupervised learning. In the k-means algorithm, first k members (k is the number of clusters) are randomly selected from n members as the centers of the clusters. The remaining n-k members are then assigned to the nearest cluster. After allocating all the members, the clusters centers are recalculated and assigned to the clusters according to the new centers. This continues as long as the centers of the clusters...
remain constant. Therefore, in k-means clustering, an objective function is optimized. The clustering responses in this method may be performed using minimization or maximization of the objective function. This means that if the criterion is the distance between objects, the objective function will be based on minimization. In fact, the clustering response is to find clusters where the distance between objects in each cluster is minimal. In contrast, if the similarity function is used to measure the similarity of objects, the objective function is selected so that the clustering response maximizes its value in each cluster. Usually, when the goal is minimization, the objective function is also called the cost function.

Suppose that the observations $(x_1, x_2, \ldots, x_n)$ that have a dimension $d$ are to be divided into $k$ parts or clusters. These clusters are known by a set called $S = \{S_1, S_2, \ldots, S_k\}$. Cluster members should be selected from observations that minimize the within-cluster sum of squares (WCSS), which is similar to variance in a one-dimensional state. Therefore, the objective function in this method is written as follows.

$$\arg\min_S \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg\min_S \sum_{i=1}^{k} |S_i| \text{Var} S_i$$

(1)

where, $\mu_i$ is the mean of $S_i$ cluster, $|S_i|$ is the number of members of cluster $i$, and $\|\|$ is the squared Euclidean distance.

It can be shown that minimizing this value means maximizing the mean squares of the distance between points in different clusters (between-cluster sum of squares: BCSS). Because according to the law of total variance, as the value of WCSS decreases, the value of BCSS increases (because the total variance remains constant).

### 2.2 PSO

The PSO algorithm is one of the nature-inspired optimization methods. This method, first developed in 1995 by Kennedy and Eberhart, is used to solve numerical optimization problems or a considerable solution space without knowing the gradient of the objective function [20]. In this algorithm, the position of each particle is a possible solution to the problem in multidimensional space. Each particle moves in the search space with a variable velocity that is dynamically modified according to the particle motion experience as well as the experience of other particle’s motion and tries to correct its position by imitating the properties of successful particles. Each particle has five properties: position, objective function corresponding to the position, velocity, best position, and objective function corresponding to the best position. The steps of this algorithm are as follows:

1. Initialization: Creating an initial population, definition of inertia weight ($\omega$) and constant coefficients $c_1$ and $c_2$.
2. Evaluation of population.
3. Determining the best personal position ($P_{best}$) and the best group position ($G_{best}$).
4. Update the velocity of each particle using the following equation:

$$v(t+1) = \omega \times v(t) + c_1 \times r_1 \times (P_{best}(t) - x(t))$$

$$+ c_2 \times r_2 \times (G_{best}(t) - x(t))$$

(2)

5. Updating the position of particles using the following equation:

$$x(t+1) = x(t) + v(t+1)$$

(3)

where, $r_1$ and $r_2$ are two random parameters within [0, 1].

6. If the stop criterion is not met, go to Step 2.
7. Return $G_{best}$ as the final solution.
8. End.

### 3 Experimental Measurements

In order to obtain the transformer TFs at different stages of the manufacturing process, several tests and measurements must be performed on the transformer. This research has performed all measurements and tests on 50 MVA, 33/132 kV oil-immersed transformers. The winding connection of the transformer is of star-delta type. The HV winding (with star connection) consists of 78 discs which each disc has eight turns. Measurements were taken at each stage for the U phase. The LV phases are grounded in all measurements. All the tests required in current research have been performed in the Iran Transfo HV laboratory. Fig. 1 shows a view of the transformer active part under test in this paper.

Fig. 2 shows the measurement circuit. As shown in Fig. 2, the input voltage ($V_{in}$) is applied to the HV terminal, and the output voltage ($V_{out}$) is measured at the HV winding common end. Therefore, the measured TF of the transformer can be defined as follows:

$$TF = \frac{V_{out}}{V_{in}}$$

(4)

Fig. 1 A view of the transformer active part under test.
The transformer must remain in the furnace for a long time, because the moisture in the transformer to be completely absorbed. Obviously, after producing the active part and before being placed in the furnace, it has the highest amount of moisture. By increasing the duration placement of the transformer in the furnace, the moisture content decreases. Therefore, relatively many measurements have been done to complete the database. The first measurement is made before the transformer is placed in the furnace. Then, the transformer was placed inside the furnace, and 24 measurements were performed at different time intervals for 72 hours. The temperature, vacuum, and amount of water lost up to the time of measurement are recorded for each measurement. The measurements taken before being placed in the furnace are considered as the reference TF, and the rest are considered the new TFs. Fig. 3 shows some of the results of the measurements. These measurements show that as the humidity in the transformer decreases, the variations in the shape of the TFs will also increase relative to the reference value. Determining the amount of moisture using the results of these measurements is a topic that will be addressed in the next section of the article.

4 Proposed Method for Moisture Determination

The main purpose of the current research is to create an intelligent pattern recognition system based on the results of TF measurements performed on a transformer so that by applying data from any other transformer to this system, its moisture content can be determined. Therefore, in the first step, the data obtained from the measured TFs must be clustered. The moisture content of a new transformer delivered to the customer is approximately 0.5% in weight. The aging of the transformer increases its water content because of the production of water due to destructive reactions in the insulating paper. When the humidity level reaches more than 5% in weight, the transformer is nearing the end of its life. According to the IEEE standard [22], a transformer with a water content of less than 2% in weight is considered dry. When the water content is between 2 and 4%, it is wet, and above 4.5%, it is excessively wet [22, 23]. According to this standard in current research, the moisture content of the transformer is classified into 3 clusters: dry, wet, and excessively wet. Table 1 shows the clustering of data based on the results of measurements.

It is important to note that clustering requires appropriate features to be extracted from the measurement results. In detecting the moisture content of the transformer using FRA, feature extraction is based on the use of information comparing the new TF with the reference TF. One of the best possible methods to compare the TFs with the reference TF is to use the NCC index [10]. This index is defined as follows:

\[
CC = \frac{\sum_{i=1}^{N} (X(i) - \bar{X})(Y(i) - \bar{Y})}{\sqrt{\sum_{i=1}^{N} [X(i) - \bar{X}]^2 \sum_{i=1}^{N} [Y(i) - \bar{Y}]^2} \\
\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X(i), \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y(i)
\]  

(5)

where, \(X\) and \(Y\) are the magnitude vector of the reference TF and the new TF, respectively. \(X(i)\) and \(Y(i)\) are \(i\)-th elements of the \(X\) and \(Y\) vectors. \(N\) is also the number of samples in a vector.

To have a sufficient amount of data, according to the Chinese standard [18], the NCC index is calculated in the following three frequency ranges:

- LF: 1-100 KHz,
- MF: 100-600 KHz,
- HF: 600-1000 KHz.

Given that 24 TFs have been measured, the feature matrix will be a 3x24 matrix. Therefore, the feature matrix can be constructed as below:

\[
\text{Input} = [\begin{array}{ccc}
NCC_{LF,1} & \cdots & NCC_{LF,24} \\
NCC_{MF,1} & \cdots & NCC_{MF,24} \\
NCC_{HF,1} & \cdots & NCC_{HF,24}
\end{array}]_{3 \times 24}
\]  

(6)

Table 1 Clustering of measurement results.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Moisture content (in percentage by weight)</th>
<th>Measurement time after being placed in the furnace (in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>More than 4</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
</tr>
<tr>
<td>2</td>
<td>Between 2 to 4</td>
<td>15, 20, 25, 30, 35, 40, 45, 50</td>
</tr>
<tr>
<td>3</td>
<td>Less than 2</td>
<td>51, 54, 57, 60, 63, 66, 69, 72</td>
</tr>
</tbody>
</table>

Fig. 2 Transformer TF measurement circuit.

Fig. 3 Some of the measured TFs of the first transformer.
where, \( i \) represents the measured TF number, which varies from 1 to 24 in this paper.

After creating the database and clustering them, the centers of the clusters must be determined using \( k \)-means. Lloyd’s algorithm is used as a standard algorithm to determine the centers of clusters [19]. However, this algorithm is very dependent on the initial position of the centers, which is determined randomly. The Lloyd algorithm falls in the local optimal points if the initial solutions are not the right choices. In this study, to solve this problem, the PSO algorithm is employed. To have optimal clustering, we minimize the sum of the distances from the centers of the clusters. Therefore, the objective function is defined as follows:

\[
\text{Cost Function: } \min \sum_{i=1}^{n} d_i \tag{7}
\]

where, \( d_i \) is the distance \( x_i \) from the nearest center \( C_j \) and is defined as follows:

\[
d_i = \min_j \| x_i - C_j \|_2 \tag{8}
\]

After finding the centers of the clusters, calculate the distance between the data obtained from the other three transformers and the centers of the clusters. Whichever has the shortest distance, the new data will belong to the same cluster. Therefore, it can be said that determining the humidity of the transformer using FRA based on the \( k \)-means-PSO algorithm consists of two phases:

**Phase 1:** Determining the centers of the clusters with the help of data obtained from a sample transformer
The steps in this phase are as follows:
1. Measuring FRA on a sample transformer and obtaining the required TFs.
2. Feature extraction with the help of (5) and (6).
3. Creating three clusters (dry, wet, and excessively wet) using the extracted features.
4. Determining the centers of the clusters by optimizing the objective function (Eq. (7)) using the PSO algorithm.

**Phase 2:** Determine the humidity of any other transformer with the help of phase 1
The steps in this phase are as follows:
1. FRA measurement on the new transformer and obtaining the required TFs.
2. Extracting the required data from the results of measurements using (5).
3. Calculating the distance of new data from the centers of the clusters using Euclidean distance (Eq. (8)).
4. Return the cluster whose center has the shortest distance from the new data.

### 5 Analysis of Results

In this section, the performance of the proposed method in detecting the moisture content of the transformer is evaluated. Part of the data is related to the first transformer, which is used to determine the centers of the clusters. The second category is data obtained from other three transformers and used to validate the proposed method.

Fig. 4 shows the data obtained from the first transformer along with its clusters centers (obtained with the help of PSO). Since the data are three-dimensional, so the data drawing and clustering are done in three states. Figs. 4(a) and (b) show respectively the variations of the NCC index in the high and medium frequency range.
frequency ranges, in terms of the variations of NCC in the low-frequency range. Fig. 4(c) shows the variations of this index in the high-frequency range in terms of its variations in the medium frequency range. As can be seen in these figures, the center of each cluster has a considerable distance from the data of other clusters. Therefore, it will be easy to decide on the humidity level of transformer insulation.

Fig. 5 also shows the convergence curve of the PSO algorithm in optimizing the cost function. As can be seen, the PSO algorithm was able to achieve the optimal solution in 40 repetitions. In the proposed method, the population size equals to 50, and the stop criterion is the number of repetitions.

After determining the centers of the clusters, the moisture content of the other three transformers is determined by phase 2. It should be noted that the characteristics and geometric dimensions of these transformers are exactly the same as the first transformer. Also, their moisture content is known, and the purpose of using these transformers is to validate the proposed method. These transformers are selected so that each belongs to one of the clusters.

Table 2 shows the results of moisture detection of these transformers.

A comparison of the transformer insulation moisture determination using the proposed k-means-PSO method with the results of the classical artificial neural network (ANN) method is also given in Table 2. The ANN used in this research is a three-layer perceptron network consisting of an input layer, a hidden layer, and an output layer. The back-propagation has been used to train the network. The hidden layer activation function is a hyperbolic tangent (known in the neural network toolbox in MATLAB software as tansig), and the output layer activation function is linear (known in the neural network toolbox in MATLAB software as purelin). As shown in Table 2, the proposed method based on k-means-PSO has correctly determined the humidity level of all transformers. However, the ANN method in one of the transformers incorrectly detected the humidity level. Therefore, it can be said that the proposed method performed better than the ANN method.

According to the excellent results obtained, the proposed method can be used to estimate the moisture content of transformers that have the exact dimensions and voltage and power levels as the transformers studied in this study. In order to generalize the proposed method, it should be tested on power transformers of different ratings.

6 Conclusion

Transformers are among the most expensive equipment in the power grid, and the guarantee of electricity supply to subscribers is closely related to their reliability. So, condition monitoring of transformers is vital. The moisture in the transformer’s insulation will reduce its life and ultimately reduce the reliability of the network. Due to the importance of this subject, in this paper, a method based on frequency response analysis using the k-means method and PSO algorithm to determine the moisture content of the transformer was proposed. By performing measurements on different transformers, their frequency response was obtained. The necessary features were extracted, and the clusters were defined by calculating the NCC index in three frequency ranges. Then, using the PSO algorithm, the centers of the clusters were determined in the k-means method. The moisture content of the new transformers was estimated by calculating the distance of new data from the centers of the clusters. The results showed that the proposed method accurately estimates the moisture level of the transformer and can be used in industry.

Intellectual Property

The authors confirm that they have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property.

Funding

No funding was received for this work.
CRediT Authorship Contribution Statement

M. Bigdeli: Idea & Conceptualization, Research & Investigation, Data Curation, Analysis, Funding Acquisition, Methodology, Project Administration, Software and Simulation, Supervision, Verification, Original Draft Preparation, Revise & Editing.

Declaration of Competing Interest

The authors hereby confirm that the submitted manuscript is an original work and has not been published so far, is not under consideration for publication by any other journal and will not be submitted to any other journal until the decision will be made by this journal. All authors have approved the manuscript and agree with its submission to “Iranian Journal of Electrical and Electronic Engineering”.

References


M. Bigdeli received the B.Sc. and M.Sc. (Hons.) degrees in Electrical Engineering from the Iran University of Science and Technology, and Zanjan University, in 2004 and 2006, respectively, and the Ph.D. (Hons.) degree in Electrical Engineering from the Islamic Azad University, Science and Research Branch, Tehran, Iran, in 2012. From 2005 to 2006, he was affiliated as a M.Sc. scholar with Iran Transformer Research Institute (ITRI). In 2006, he joined Islamic Azad University, Zanjan Branch, where he is now Assistant Professor. From 2019, he has also been working with Zanjan Regional Electricity Company and Iran Transfo Company as a consultant and project manager. His main research interests include fault detection, transient modeling, and optimization of electrical machines, especially transformers, renewable energy, heuristic algorithms, and pattern recognition methods. In these research fields, he has co-authored more than 50 articles, including more than 40 international journal papers that received more than 450 citations in Scopus with an H-index equal to 10.

© 2022 by the authors. Licensee IUST, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (https://creativecommons.org/licenses/by-nc/4.0/).