Characterization of the operating periods of a power transformer by clustering the dissolved gas data

Conference Paper · August 2017

DOI: 10.1109/DEMPED.2017.8062371

4 authors, including:

Samuel Eke
University of Douala, Faculty of Industriial Engineering, Douala Cameroon

Issouf Fofana
University of Québec in Chicoutimi

Characterization of the Operating Periods of a Power Transformer by Clustering the Dissolved Gas Data

Samuel EKE
Univ. Lyon – Ampère (CNRS UMR 5005), Ecole Centrale Lyon, Univ. Claude Bernard Lyon 1, INSA de Lyon, France
samuel.eke@doctorant.ec-lyon

Guy CLERC
Univ. Lyon – Ampère (CNRS UMR 5005), Ecole Centrale Lyon, Univ. Claude Bernard Lyon 1, INSA de Lyon, France

Thomas AKA-NGNUI
Univ. Lyon – Ampère (CNRS UMR 5005), Ecole Centrale Lyon, Univ. Claude Bernard Lyon 1, INSA de Lyon, France

Issouf FOFANA
Research Chair on the Aging of Power Network Infrastructure(ViAHT)
Univ. du Québec à Chicoutimi line 3-City, Country
QC G7H 2B1, Canada

***Abstract*— This paper presents an analysis of the different operating periods of an in-service oil immersed power transformer through dissolved gas concentrations. The unsupervised classification by k-means method allows regrouping the periods of operation into classes using the Euclidean distance as a criterion of similarity. The analyzed data describes the evolution of gas concentrations as a function of time. The classes obtained are characterized by the production activities of the different gases, various operating constraints and the incipient failures. These periods also highlight the maintenance actions carried out on the insulating oil.**

Keywords*— *Data analysis, Diagnostic, Dissolved gas analysis, K-means, Predictive maintenance, Principal Component Analysis, Transformers, Machine learning, Unsupervised Learning.

I. INTRODUCTION

POWER transformers play an essential role in the supply of electric power. The consequences of their unavailability lead to enormous losses for both the network operators and for end-users. In order to limit the risk of their unavailability, accurate maintenance procedures must be considered. Among the actions to be undertaken, the monitoring of their insulation is widely considered. Oil is regulary sampled from the tank and On-Load-Tap-Changer (OLTC) allows monitoring the condition of the unit. From the physico-chemical analyses of these samples, the quality of the oil, and the ability to play its role effectively in the transformer are assessed. The ASTM D3612 standard describes procedures to extract the dissolved gases from the oil [1],[2]. Identification and concentration of each gas can be determined by gas chromatography. The considered gases dissolved in the oil are: hydrogen (H_2), methane (CH_4), ethylene (C_2H_4) ethane (C_2H_6), acetylene (C_2H_2) , carbon monoxide (CO), carbon dioxide (CO_2), oxygen (O_2) and nitrogen (N_2). These gases are the outcome of a chemical decomposition of the oil and the cellulose insulation under incipient electrical and thermal failures.

The Dissolved Gas Analysis (DGA) is widely recognized as a reliable method for evaluating the internal condition of the transformer and diagnosing incipient faults. These recent works illustrate it very well [3], [4], [5]. The concentration of these gases, in particular the key gases, are used for a quick diagnosis of certain thermal and electrical faults. Standards [6], [2] describe the mechanism for the formation of key gases and their causes under certain conditions. These causes can be classified into four categories:

- Partial discharges and corona: these failures are accompanied by a large production of hydrogen H_2 and a low production of methane CH_4. These faults are also known as low energy discharges;
- The second category is related to the oil degradation. It is characterized by the generation of ethylene C_2H_4, methane CH_4 and ethane C_2H_6.
- The third category consists of the formation of arcs and sparks with a high energy release. This is associated to a large amount of hydrogen H_2 and acetylene C_2H_2.
- The fourth category is related the thermal degradation of cellulose. It is characterized by the formation of a large amount of carbon monoxide CO (combustible gas).

Fig. 1, reproduced from [7], [8] graphically sketches these decomposition mechanisms.

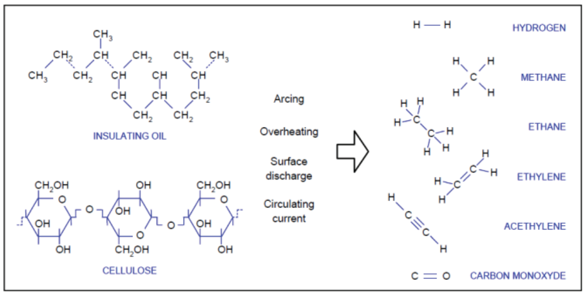

Fig. 1.Key gases production illustration [7], [8].

Several methods have been reported in the literature for the interpretation of the dissolved gases [9], [10], [11], [12], [13],[14]. The Total Dissolved Combustible Gas (TDCG) is the sum of the concentration of all combustible gas components (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2 and CO). It is one of the transformer condition indicators. Ref [14] reports four criteria to classify the risks in transformers based on the values of combustible gases.

978-1-5090-0409-6/17/$31.00 ©2017 IEEE

This study takes particularly into account key gases evolution. These gases are used to identify the operating periods of a transformer based on its maintenance data. These periods are pre-obtained by clustering with the k-means method. This contribution directed towards the reconstructions of the most significant periods of the considered transformer's life. The information are extracted automatically and used for the prediction of potential incipient fault and appropriate maintenance actions proposed.

The methodology applied in this study is in the context of machine learning.

This paper is subdivided in seven sections:

- The characteristics and data of the transformer studied are presented in section II after an introduction in section I.
- Section III presents the methodology used to extract knowledge from 16 years gas concentration data.
- The results are presented in section IV.
- Section IV presents the class characterization.
- In section VI, a discussion of the results is exposed.
- Section VII presents a conclusion and suggests some ideas for future work.

II. TRANSFORMER CONSIDERED IN THIS STUDY

A. Power Transformer characteristics

Table I presents the characteristics of the transformer under study. It is a Generator Step-Up (GSU) transformer without an On-Load-Tap-Changer.

TABLE I
CHARACTERISTICS OF TRANSFORMER

MANUFACTURER : CGE ; SERIAL N° : 270074	
VOLTAGE	24 KV
POWER	150000KVA
YEAR OF COMMISIONING	1952
SITE OF OPERATION	CENTRALE ISLE MALIGNE (CIM) ON THE SAGUENAY RIVER IN CANADA
EXPLOITATION	RIO TINTO ALCAN CANADA
DATA	1992 - 2008

B. Data

The maintenance data for the transformer, which has the characteristics reported in Table I, cover all the concentrations of the different gases resulting from the gas chromatography analysis. From 1992 to 2008, the insulating oil of this transformer was analyzed and the concentrations of dissolved gases were recorded. The initialization of the observation is done from 1992 and each observation becomes a cumulative duration. These data are in the form of a matrix with n = 42 observations and p = 6 variables or attributes. In other words, each row of the matrix is the cumulative duration and the corresponding columns are the gas concentrations (H_2, CH_4, C_2H_4, C_2H_6, C_2H_2, and CO).

III. METHODOLOGY

The objective is to automatically extract knowledge from the dissolved gas concentrations data for a transformer tracked over time. More or less, unsuspected relationships that carry information about the behavior of this transformer are sought. This information may be: fault under the effect of operating stresses, or any other action enabling to understand the past service of this transformer. First, a Principal Component Analysis (PCA) is applied to identify the variables that will serve as 3D representation. Then, a clustering is performed as described in [15] by the k-means classification method. The characterization of the obtained classes is performed by the analysis of the Total Dissolved Combustible Gas (TDCG) and each key gas.

A. Determination of data representation space by PCA

The PCA is used here in pre-processing of the k-means classification, to represent each data with a reduced number of variables. This gives a compact representation of the data. At the same time, the information contained in the cluster representation space is quantified.

Let $\mathbf{X} \in R^{n\times p}$ be the matrix containing n data (observations). Each observation is described by p variables. The data is considered as a point in a p dimensional space.

The objective of the PCA is to construct the Euclidean space of the most characteristic and economic dimensions for data representation. With the PCA, it is possible to move from the data space to the factorial space. This space is obtained by axes, which represents the main components, i.e. the linear combinations of the initial data.

The correlation matrix $\mathbf{R}$ is obtained by this equation [16]:

$$\mathbf{R} = \frac{1}{p}\mathbf{X}^t\mathbf{X} \tag{1}$$

where $\mathbf{X}^t$ is the transpose matrix of $\mathbf{X}$.

The principal components are obtained by calculating the eigenvector and eigenvalues of $\mathbf{R}$. They are identified by the largest eigenvalues of the matrix $\mathbf{R}$ and the associated eigenvectors. The relative importance of an eigenvalue is measured by its inertia, and is defined by equation (2) [16].

$$I_i = \frac{\lambda_i}{p} \tag{2}$$

where, $\lambda_{i\in\{1,\dots,p\}}$ are the eigenvalues of the matrix $\mathbf{R}$.

B. K-means theory

Clustering consists in grouping data in an unsupervised way without the contribution of an expert. The objective is to automatically extract information from the data. Data of the same group are called cluster and are closer to each other than those of other clusters, in the sense of a criterion of (dis) similarity. In other words, each data is assigned to a cluster, if it is very close to its center of gravity.

Consider $(x_1,\dots,x_n)$ a set of data. Let's define $\mathbf{\Gamma}$ as the classification distance criterion, which evaluates the distance of a given data to the centers of the formed classes. Let

$C=(c_1,\ldots,c_K)$ be the set of class centers and $G=(g_1,\ldots,g_n)$ the cluster set.

Γ, considered in the study is the total Euclidean distance between each data and the center of the closest class.

$$\Gamma(C,G)=\sum_{k=1}^{K}\sum_{i=1}^{n} g_{ik}\left\|x_i-c_k\right\|^2 \quad (3)$$

The Euclidean distance is defined by:

$$d(x_i,c_k)=\left\|x_i-c_k\right\|^2=\sqrt{\sum_{j=1}^{d}(x_{ij}-c_k)^2} \quad (4)$$

In equation (3), g_{ik} is a binary variable equal to 1 if the cluster of the indexed data is k and 0 otherwise. The K-means algorithm [15],[17] can be summarized in three steps.

1- Initialization:

The class centers are initialized $(c_1^{(0)},\ldots,c_K^{(0)})$ by fixing a random value of K. In other words, it is the random choice of virtual centers, to start the iteration at t = 0.

2- Clustering step:

Each data is assigned to the class of the closest center.

$$\forall i=1,..n\,,\, g_{ik}^{(t)}=\begin{cases}1 \text{ if } k=\arg\min\limits_{g\in\{1,..K\}}\left\|x_i-c_g\right\|^2\\ 0 \text{ other wise}\end{cases} \quad (5)$$

3- Step of recalibration of the centers

$$\forall i=1,..k\,,\; c_k^{(t+1)}=\frac{\sum_{i=1}^{n} g_{ik}^{(t)}x_i}{\sum_{i=1}^{n} g_{ik}^{(t)}} \quad (6)$$

where t is the current iteration.

Convergence can be considered achieved if the relative value of criterion Γ falls below a prefixed small threshold or if a maximum number of prefixed iterations have been reached. Once the algorithm has converged, the estimation of the classes $(g_1,\ldots,g_n)$ and their centers $(c_1,\ldots,c_K)$ is obtained. Thus, it is possible to correlate each data to a cluster using certain evaluation criteria of the clustering quality as Dun Index [18], Silhouette Index [19] or Davies-Bouldin Index [20].

IV. Results

These results reported hereafter are obtained from the transformer data presented in Section II. The feature vector consists of the combustible gases H_2, CH_4, C_2H_4, C2H6, C_2H_2, and CO.

A. PCA

Table II presents the eigenvalues corresponding to the main components, and cumulative variances, which highlight the information carried by the space representation. The table is derived from the calculation of correlation matrix $\mathbf{R}$.

TABLE II
PRINCIPAL AXES, EIGENVALUES, EXPLAINED VARIANCE AND CUMULATIVE VARIANCES

Factor	Eigenvalues of R	Explained variances (%)	Cumulated variances (%)
1	3.2803	54.6712	54.6712
2	1.3310	22.1830	76.8542
3	1.0647	17.7450	94.5992
4	0.2153	3.5881	98.1873
5	0.0686	1.1440	99.3313
6	0.0401	0.6687	100

The first three factors account for 94.59% of the information in the data set (cumulated variances in Table II).

The projection of the initial variables in the space constituted by these three main factors allows identifying the information carried by each factor, and thus characterizes the new variables. The coordinates of these variables are reported in Table III. Fig. 2 shows the correlation circle, projection of the variables on the first two axes of Table III. This Table is obtained by considering the eigenvectors of $\mathbf{R}$.

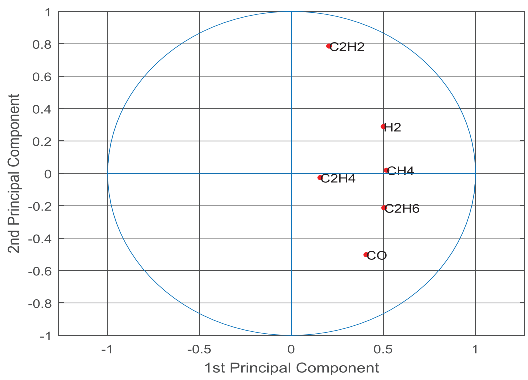

Fig. 2. Correlation circle

TABLE III
CORRELATION BETWEEN AXES AND VARIABLES

Variables	Axe1	Axe2	Axe3	Axe4	Axe5	Axe6
H2	0.4997	0.2890	-0.1345	-0.3569	0.2736	0.6681
CH4	0.5163	0.0197	-0.1044	0.6639	-0.5095	0.1476
C2H2	0.2037	0.7862	0.1369	-0.2017	-0.1967	-0.4921
C2H4	0.1559	-0.0261	0.9216	0.2135	0.2705	0.0836
C2H6	0.5032	-0.2124	-0.2561	0.1478	0.6010	-0.5031
CO	0.4060	-0.5021	0.1933	-0.5689	-0.4388	-0.1717

The axis 1 groups the information carried by the variables CH_4 and C_2H_6. It is also possible to associate the information carried by H_2 since only three factors are considered. The axis 2 carries the information of the variable C_2H_2, and the opposite to those of the variable CO. The axis 3 carries the information of the variable C_2H_4.

The PCA allowed identifying the new variables that will constitute the data visualization space. Each axis selected thus constitutes a new variable of the representation space.

B. K-means analysis

Figure 3, shows the projection of the data in the three dimensional space described by the variables that best summarize the information contained in the data. The Silhouette index and that of Davies [3] allowed identifying four optimal clusters. A data structure has thus been obtained in four clusters.

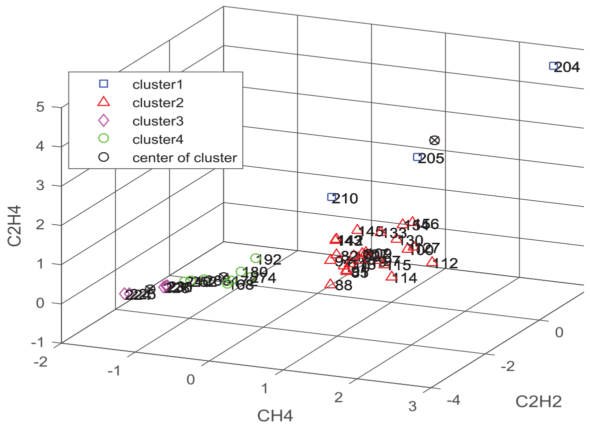

Fig. 3. Clustering of the data by K-means.

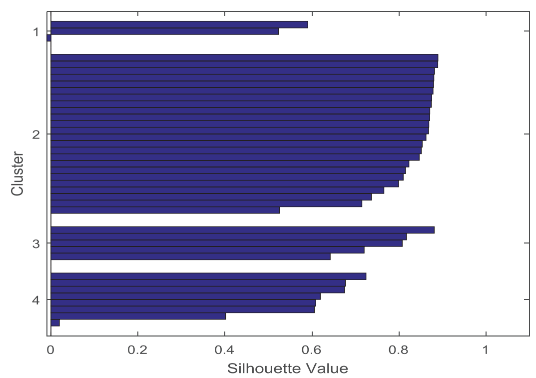

Fig. 4. Cluster evaluation by diagram silhouette.

A quick analysis of the formed clusters using the silhouette plot (Fig. 4) allows noting the following:

- In the first cluster, there are 3 elements and one of them is badly classified.

- In the second cluster, there are 24 well-classified elements.

- The third cluster consists of 5 elements and the fourth cluster contains 12 elements, one may not be misclassified.

In Fig. 5 and Fig. 6, the classes are chronologically represented. By superimposing them on the TDCG and the concentrations of each gas, the activities of these gases during the periods described by the classes are observed.

V. Class Characterisation

Data used to make the classification are dissolved gases in the transformer insulating oil concentrations. The DGA helps diagnosing incipient faults in the transformer [6]-[14]. These faults can be caused by thermal (various heating) or electrical stresses (partial discharges and high-energy arcing). The periods determined by the classification reveal particular activities of gas production in the oil. The incipient faults are then identified by following key gases and the TDCG evolution.

The classification by k-means provides four clusters:

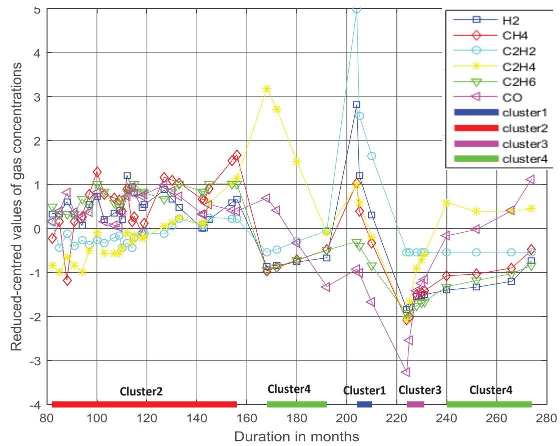

Fig. 5 Clusters and all gases time evolution

- Cluster 1 is the period, which regroups the duration 204, 205 and 210. This period is characterized by increasing values of acetylene (C_2H_2), between 9 ppm and 16 ppm, which largely exceed the limit value (1 ppm, fig. 7). The concentration of hydrogen (H_2) is remarkable but does not exceed the limit value of 100 ppm (fig. 8).

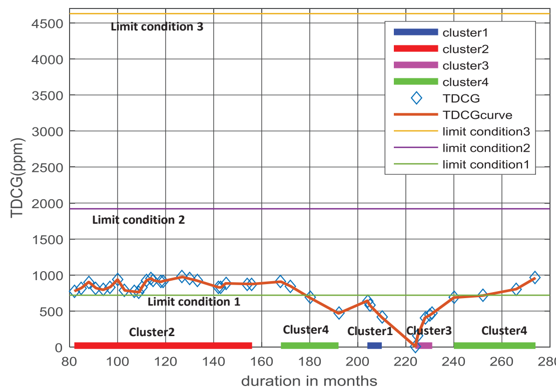

Fig. 6. Clusters and TDCG time evolution

According to authors in [21], high values of C_2H_2 and H_2 without necessarily high TDCG value, indicates the existence of an electric arc with production of a large amount of energy. An oil treatment or replacement is recommended and an inspection of the internal parts of the transformer must be performed. It can be observed on the TDCG curve of Fig. 6

that after this period, the TDCG dropped to a value of 6 ppm and the resumption of the TDCG increases immediately after, confirming that an action has been taken on the oil.

- Cluster 2 illustrates the operation of the transformer with TDCG values between 760 ppm and 1000 ppm. It is classified in condition 2 (Fig. 6). This period is also marked by CO values, which largely exceed the limit value of 350 ppm, indicating an overheating.

- Cluster 3 is characterized by a rise in concentration of all gases (fig. 5).This can be explained by the fact that the oil was certainly treated but all the problems of the transformer were not detected. An analysis of the carbon monoxide (CO) values indicates that the limit value (350 ppm) have been exceeded, which indicates a heating, a hot spot or a connection problem in the tank [6]. This period is marked by an important production of CO with an increase of all key gases. However, this increase is less important in Cluster 4.

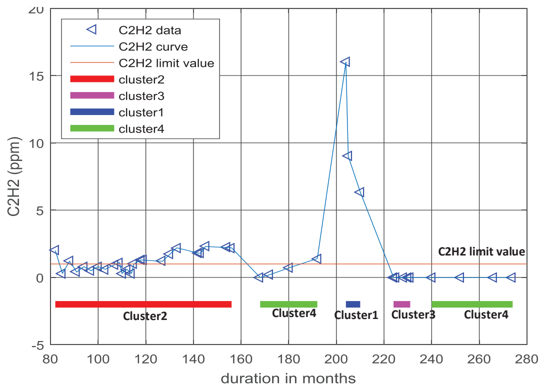

Fig. 7 Clusters and C_2H_2 time evolution

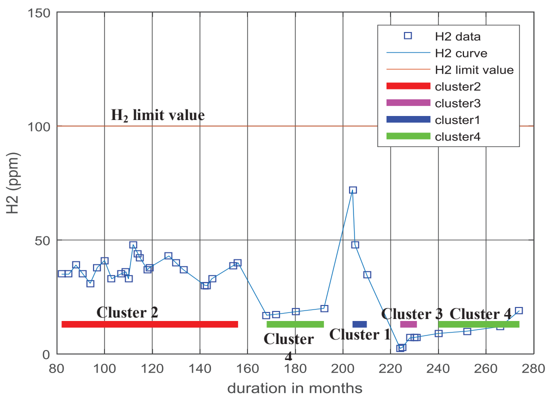

Fig. 8 Clusters and H_2 time evolution

- Cluster 4 is characterized by high values of C_2H_4 and increasing values in other gases, except for CO which decreases in the first part of cluster 4 and increases in the second part. This period corresponds to a post treatment of the oil. It should be noticed that the preceding period is marked by a temperature rise with CO values that exceed the limit (Fig. 9). The second part of this class confirms the trends of the first one: post-oil treatment period with the same problem of excessive CO values. This transformer has a hot spot problem on a connection to the tank.

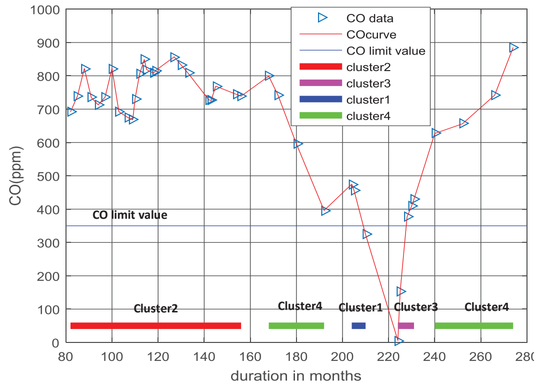

Fig. 9. Clusters and CO time evolution

VI. Discussion

The analysis proposed in this study allows identifying the significant operating periods of the power transformer under consideration. It is based on the dissolved gases data in the oil produced under various operating conditions of the transformer. It is quite possible to have other events that would be noted if other data are included in the analysis. For example, others components coming from the same conditions such as furan contents, particles in oil, presence of copper etc.. However, although the use of k-means gives coherent results, it is not impossible that another method such as the fuzzy k-means or the HAC (Hierarchical Ascending Classification) produces the same results or improves the classes achievement . Nevertheless, the obtained classifications using k-means method are consistent and allow understanding and summarizing the significant operating periods of this transformer.

VII. Conclusion and Future Work

The objective of this paper is to automatically extract the groups (classes) in the dissolved gases data in the insulating oil of the studied transformer. The interpretation of this classification was based on the information provided by each gas and the TDCG in terms of fault and alert for possible maintenance actions. The characterization of these classes led to the identification of major events (faults) that marked each class. The first class revealed the presence of an electric arcing, with high energy. The second cluster revealed an abnormal temperature rise of the oil. The third cluster presented a period characterized by an accelerated increase in the production of all gases. As for the fourth cluster, it allowed to notice the post treatment periods of the oil.

It will be interesting in the future to investigate the prediction of the transformer condition by following the evolution of dissolved gases. It will be interesting to apply supervised classification methods to extract association rules from the Standard [14].

ACKNOWLEDGMENT

The authors gratefully acknowledge the cooperation of Rio Tinto in Canada, for providing maintenance data for transformers at their power plant. This study is based on one of the GSU transformers of the Isle-Maligne plant. The authors also thank the University of Douala in Cameroon and the Faculty of Industrial Engineering for the funding of the thesis under which this work is carried out.

REFERENCES

[1] Nick Lelekakis, Daniel Martin, Wenyu Guo and Jaury Wijaya, “Comparison of Dissolved Gas-in-Oil Analysis Methods Using a Dissolved Gas-in-Oil Standard” *IEEE Trans. Electrical Insulation*, vol. 27, number 5, pp. 29-35, Oct. 2011.
[2] ASTM: Analysis of Gases Dissolved in Electrical Insulating Oil by Gas Chromatography*, ASTM D 3612-02, 2009.*
[3] J. Faiz and M. Soleimani, "Dissolved gas analysis evaluation in electric power transformers using conventional methods a review," in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 2, pp. 1239-1248, April 2017.
[4] E. Wannapring, C. Suwanasri and T. Suwanasri, "Dissolved Gas Analysis methods for distribution transformers," *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Mai, 2016, pp. 1-6.
[5] T. Suwanasri, P. Sirimongkol, C. Suwanasri and R. Phadungthin, "Software development for power transformer condition evaluation using dissolved gas analysis methods; practical experience in Thailand," *2015 18th International Conference on Electrical Machines and Systems (ICEMS)*, Pattaya, 2015, pp. 329-333.
[6] EN 60599:Matériels électriques remplis d'huile minérale en service- Lignes directrices pour l'interprétation de l'analyse des gaz dissous et des gaz libres, NF EN 60599,C27-239, Avril 2016
[7] E MENTLÍK, V. et al. *Spolehlivostní aspekty elektrotechnologie*.1. vyd. Praha: BEN - technická literatura, 2011, 118 p.
[8] Jan DONČUK, Václav MENTLÍK, Jiří VELEK**,** *Condition assessment of power transformer using gas detection Methods***,** PRZEGLĄD ELEKTROTECHNICZNY, 2013.
[9] M. Duval, “State-of-the-art of technical diagnosis through interpretation of dissolved gas analysis” in *Proceedings of My Transformer Conference*, Torino, Nov. 21–22, 2012.
[10] M. Duval and L. Lamarre, "The duval pentagon-a new complementary tool for the interpretation of dissolved gas analysis in transformers," in *IEEE Electrical Insulation Magazine*, vol. 30, no. 6, pp. 9-12, November-December 2014.
[11] Lewand, L. “Techniques for Interpretation of Data for DGA from Transformers”, in Proceedings of IEEE/PES Transformers Committee Meeting, Montreal, QC, Canada, 26 October 2006.
[12] I. B. M. Taha, H. G. Zaini and S. S. M. Ghoneim, "Comparative study between dorneneburg and rogers methods for transformer fault diagnosis based on dissolved gas analysis using Matlab Simulink Tools," *2015 IEEE Conference on Energy Conversion (CENCON)*, Johor Bahru, 2015, pp. 363-367.
[13] Sungjik Kim, Hwangdong Seo and Jaeryong Jung, "Advanced dissolved gas analysis method with stray gassing diagnosis," *2016 International Conference on Condition Monitoring and Diagnosis (CMD)*, Xi'an, 2016, pp. 522-525.
[14] IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, IEEE Standard C57.104-2009, Feb. 2009.
[15] Anil K. Jain, Richard C. Dubes, *Algorithms for clustering*, Prentice Hall advanced Reference Series, Englewood Cliffs New Jersey, 1948, pp. 55-141.
[16] Brigitte Escofier, Jérôme Pagès, *Analyse factorielles simples et multiples*, Cours et études de cas, 5ème édition, Dunod, 2016.
[17] Maha Ghribi, Pascal Cuxac, Jean-Charles Lamirel, Alain Lelu, Mesures de qualité de clustering de documents : Prise en compte de la distribution des mots clés. Conférence internationale Francophone sur l’extraction et la gestion des connaissances, ECG2010, Tunisie, Janvier 2010, [Online], Available: *https*://*hal*.archives-ouvertes.*fr*/*hal-00614071*.
[18] Dunn JC (1974) *Well-separated clusters and optimal fuzzy partitions*. Journal of cybernetics 4: 95–104.
[19] Peter J. ROUSSEUW(1987, Nov) *Silhhouette: a graphical aid to the interpretation and validation of cluster analysis*, Journal of Computational and Applied Mathematics 20(1987)53-65 North Holland.
[20] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227, April 1979.
[21] N.K. Sharma, Tiwari, P.K. Sood, Yog Raj(2011, Aug.) Review of Artificial Intelligence Techniques Application to Dissolved Gas Analysis on Power Transformer, *International Journal of Computer and Electrical Engineering, Vol. 3, No. 4.*