Application of affinity propagation algorithm based on manifold distance for transformer PD pattern recognition

B G Wei¹, K X Huoº, Z F Yao¹, J Lou² and X Y Li²

¹ Electric Power Research Institute, Shanghai Electric Power Company, No.171 Handan Road, Shanghai, China.
² School of Electrical Engineering, Shandong University, No.17923 Road, Jinan, Shandong Province, China.

E-mail: wbgsj@126.com

Abstract. It is one of the difficult problems encountered in the research of condition maintenance technology of transformers to recognize partial discharge (PD) pattern. According to the main physical characteristics of PD, three models of oil-paper insulation defects were set up in laboratory to study the PD of transformers, and phase resolved partial discharge (PRPD) was constructed. By using least square method, the grey-scale images of PRPD were constructed and features of each grey-scale image were 28 box dimensions and 28 information dimensions. Affinity propagation algorithm based on manifold distance (AP-MD) for transformers PD pattern recognition was established, and the data of box dimension and information dimension were clustered based on AP-MD. Study shows that clustering result of AP-MD is better than the results of affinity propagation (AP), k-means and fuzzy c-means (FCM). By choosing different k values of k-nearest neighbor, we find clustering accuracy of AP-MD falls when k value is larger or smaller, and the optimal k value depends on sample size.

1. Introduction

Power transformer is one of the more important equipment in the power grid. If a fault occurs, power transformer can cause that partial or even large area blackouts and great economic losses. A large number of fault statistics show that most of the accidents of transformers are caused by aging and damage of insulation, and PD is one of the important reasons for the aging and damage of insulation. It has a great effect on judging the damage condition of insulation and finding the insulation defects of transformers that the types of PD are effectively identified.

A great deal of research work has been done on the diagnosis and identification of PD defects in power transformer at home and abroad [1-3]. Fingerprint recognition technology based on PD statistical spectrum has received extensive attention, and scholars have proposed PRPD, phase resolved pulse sequence analysis (PRPSA) and time-resolved partial discharge (TRPD) model based on different measurement methods and statistical features of PD. With the development of computer and mathematical methods, a variety of pattern recognition methods have emerged, such as pattern recognition method based on clustering analysis[4], pattern recognition method based on artificial neural network [5], pattern recognition method based on fractal theory [6].

As an unsupervised analysis method, clustering analysis has been widely used in pattern recognition. Krivda [7], of Delft University of Technology in Holland, used cluster analysis and artificial neural networks to discuss the automatic recognition of local scenes. Chen [8] identified the
developing stages of air-gap discharge in oil-paper insulation based on cluster-wavelet neural network. Since Frey and Dueck put forward Affinity Propagation (AP) in Science magazine in 2007 [9], AP has been widely used in image recognition, image retrieval, data mining and other fields. In this paper, we propose AP-MD and recognize three kinds of PD models by AP-MD.

2. AP algorithm

AP algorithm is a kind of clustering algorithm based on passing messages between adjacent points. In contrast to k-means and FCM, AP uses the similarities between data points as the input. First, all data points are regarded as potential category center with equal status. Then, the nearest central points are the optimal representative points of all categories when the sum of the similarities between data points and the nearest central point is the largest [10].

The data set X={x1, x2, …, xN} contains the category set C={C1, C2, …, Ck}, K ≤ N. Each data point only belongs to one category. The central point of category which data xi belongs to is xC(i), i=1,2,...,K. The error function of clustering is Equation (1).

$$J(C) = \sum_{i=1}^{N} d^2(x_i, x_{C(i)})$$  \hspace{1cm} (1)

The objective function of the AP algorithm is to minimize the error function:

$$C^* = \arg\min[J(C)]$$  \hspace{1cm} (2)

The similarity matrix S_{N×N} is calculated by Equation (3), the non-diagonal element s(i,j) of the matrix is the similarity between the points x_i and x_j, the diagonal element s(i,i) is the preferences p(i) which is negative. The initial values of p(i) generally are the same value that is the minimum or mean of all non-diagonal elements in the similarity matrix.

$$s(i,j) = \begin{cases} -\|x_i - x_j\|^2 & i \neq j \\ p & i = j \end{cases}$$  \hspace{1cm} (3)

The core of AP algorithm is passing messages between data points. The messages of the AP algorithm are ‘Responsibility’ and ‘Availability’. Responsibility r(i,j) is sent from data point i to candidate exemplar point j and it reflects the accumulated evidence for how well-suited point j is to serve as the exemplar for point i. Availability a(i,j) is sent from candidate exemplars points j to point i and it reflects the accumulated evidence for how appropriate it would be for point i to choose point j as its exemplar. The iterative process of the AP algorithm is the process of alternately updating the messages. At the beginning, the responsibilities and the availabilities are initialized to zero: r(i,j) = 0, a(i,j) = 0. The iterative equations of r(i,j) and a(i,j) are written as Equation (4) and Equation (5).

$$r(i,j) \leftarrow s(i,j) - \max_{j \neq j} \left[ a(i,j) + s(i,j) \right]$$  \hspace{1cm} (4)

$$a(i,j) \leftarrow \begin{cases} \min_{i \neq j} [0, r(i,j) + \sum_{i \neq j} \max[0, r(i,j)], i \neq j] \\ \sum_{i \neq j} \max[0, r(i,j)], i = j \end{cases}$$  \hspace{1cm} (5)

Finally, after a large number of iterations, find the optimal central points of categories and the subordinate relationship between the central points and the data points by the objective function Equation (2).

3. AP-MD algorithm

As we can see from Equation (3), the similarity matrix S_{N×N} of AP is calculated based on Euclidean distance between two data points. It makes S_{N×N} can only reflect the local consistency of data, and can’t reflect the global consistency and the potential complex structure of data. As shown in Figure 1, the Euclidean distance between point a and point b is longer than the Euclidean distance between a and b, which can cause the similarity between b and a is less than the similarity between c and a, point b and point a can be wrongly classified as one category.
3.1. Manifold distance

Before defining the manifold distance, the length of the line between two data points in manifold is defined. The definition is as follows, search the k-nearest neighbor of each sample point, and then construct the weighted graph \( G(V, E) \) by the k-nearest neighbor. \( V \) is the set of vertices that correspond to sample points, \( E \) is the set of edges. \( L(x_i, x_j) \) represents the distance between point \( x_i \) and point \( x_j \). The calculation equation is Equation (6).

\[
L(x_i, x_j) = \begin{cases} 
1 - \exp[-\beta d(x_i, x_j)^2] & x_i \text{ and } x_j \text{ are } k\text{-nearest neighbor of each other} \\
\infty & \text{else}
\end{cases}
\]  

(6)

In Equation (6), \( d(x_i, x_j) \) is the Euclidean distance between two points; \( \beta \) is a regulatory factor to prevent the rapid growth of \( L(x_i, x_j) \). The value of \( \beta \) is related to the density of the data and it’s usually the reciprocal of the mean Euclidean distance of all sample points.

The data points are considered as the vertexes of the weighted graph \( G(V, E) \). \( p = \{p_1, p_2, \ldots, p_l\} \in V \) represents a path between points \( x_i \) and \( x_j \) and the edges \( \{p_k, p_{k+1}\} \in E, 1 \leq k \leq l-1 \). \( P_{ij} \) represents the set of all paths between the points \( x_i \) and \( x_j \). The manifold distance between point \( x_i \) and point \( x_j \) is defined as Equation (7) [11].

\[
D(x_i, x_j) = \min_{p=p_i} \sum_{k=1}^{l-1} L(p_k, p_{k+1})
\]  

(7)

3.2. The steps of AP-MD algorithm

The steps of AP-MD algorithm are as follows:

- **Input**: \( k \) of k-nearest neighbor and preference \( p \).
- **Initialization**: \( r(i, j) = 0 \), \( a(i, j) = 0 \) for all \( i, j \).
- \( \forall x_i, x_j \in X \), calculate the Euclidean distance between point \( x_i \) and point \( x_j \) by Equation (8).

\[
d(x_i, x_j) = \left\| x_i - x_j \right\|
\]  

(8)

- Search the k-nearest neighbor of each sample point based on Euclidean distance, and then construct the weighted graph \( G(V, E) \).
- \( \forall x_i, x_j \in X \), calculate the manifold distance between point \( x_i \) and point \( x_j \) by Equation (7).
- Construct similarity matrix \( S_{Non} \) based on manifold distance. The non-diagonal element \( s(i, j) = d_L(x_i, x_j)^2 \), the diagonal element \( s(i, i) \) is the mean of all non-diagonal elements in the similarity matrix \( S_{Non} \).
- Use AP algorithm to cluster data. Take Equation (2) as the objective function, and perform iterative computations by Equation (4) and Equation (5).
- Determine whether the number of categories satisfies the requirement. If not satisfied, change the value of \( p \) and repeat the whole iterative process until the number of categories meets the requirement, and output the final clustering results.
3.3. Comparison of AP-MD with other algorithms

Besides the AP algorithm, the k-means algorithm and FCM are two commonly used clustering algorithms, and they both have good results in clustering. We use AP-MD, AP, k-mean and FCM to cluster the data in the Figure 1. The results are shown in Figure 2.

![Figure 2. Clustering results of four algorithms.](image)

Figure 2 shows clearly that the clustering result of AP-MD is better than the results of AP, k-means and FCM and the accuracy of the recognition reaches 100%. Clustering results of AP, k-means and FCM are only different in the black square. AP, k-means and FCM use Euclidean distance to compute the similarity matrix $S_{nch}$. So their clustering results are not only very similar, but also have the same problem which the clustering results are completely wrong. The clustering result of AP-MD is better than results of other algorithms for data which are distributed regularly.

3.4. The effects of k value on AP-MD

According to Equation (6) and Equation (7), k value of k-nearest neighbor is the key for computing manifold distance. In order to study effects of k value on clustering results of AP-MD, values of k are all integers between 1 and 50 and we cluster the data of Figure 1 by AP-MD. Clustering results are shown in Figure 3 when k values are 4, 10, 28 and 30. When $1 \leq k \leq 4$ and $27 \leq k \leq 29$, the accuracy of clustering is between $50\%$ and $100\%$. When $5 \leq k \leq 26$, the accuracy of clustering can reach up to $100\%$. Clustering results of AP-MD are similar with clustering results of AP when $k \geq 30$ and they are wrong.

![Figure 3. Clustering results of k = 4, 10, 28 and 30.](image)

According to the steps of AP-MD algorithm, it needs iterations and adjusting preference $p$ to make number of categories meets the requirement. As shown in Table 1, with the increase of k value from 5 to 26, the number of iterations is reduced and $p$ value is increased. The number of iterations reduces rapidly when $5 \leq k \leq 15$, and it reduces slowly as the k value continues to increase. Considering all effects of k value on AP-MD, The optimal k value is $10\% \sim 15\%$ of the amount of data contained in one category and doesn’t be less than 4.

| $k$  | The number of iterations | preference $p$   |
|------|--------------------------|-----------------|
| 5    | 55                       | $-2.26 \times 10^4$ |
| 10   | 48                       | $-6.32 \times 10^3$ |
| 15   | 42                       | $-2.11 \times 10^3$ |

Table 1. The effects of k value on iterations and preference $p$. 

[1] ICMES 2017 IOP Publishing
IOP Conf. Series: Materials Science and Engineering 339 (2018) 012014 doi:10.1088/1757-899X/339/1/012014
20 & 41 & -1.76x10^3 \\ 25 & 39 & -1.22x10^3 \\ \hline

4. Application of AP-MD for PD Pattern recognition

4.1. Model experiment of PD in transformers

According to the main physical characteristics of PDs in the transformers, three oil-paper insulation defects were designed which were used to simulate three typical PDs in the transformers. As shown in Figure 4. All models were immersed in oil during the experiment. The diameter of circular plate electrodes is 80mm, and the thickness is 10mm. The thickness of all paperboard is 0.5mm. Figure 4 (a) is the cavity discharge model inside solid insulation. The cavity consists of three layers of paperboard with a diameter of 60mm, and the center of the middle paperboard has a circular hole with a diameter of 20mm. Figure 4 (b) is surface discharge in oils model. The sample is a paperboard of 100mm in diameter. Figure 4 (d) is corona discharge in oils mode. The radius of curvature of the tip is less than 0.1mm. Paperboard is placed between the tip and the plate electrode, and the d is 1mm [12].

![Figure 4](image)

\textbf{Figure 4.} Three oil-paper insulation defects: (a) cavity discharge, (b) surface discharge, (c) corona discharge.

4.2. Constructing the grey-scale images

Use sampled data from the experiment to contract grey-scale images. Steps are as follows. First, extract the PD pulse, and then construct PRPD and space curved face of $\phi$-$q$-$n$ ($\phi$ is power frequency phase angle of discharge, $q$ is discharge amplitudes, $n$ is discharge times). Finally, construct grey-scale images of space curved face of $\phi$-$q$-$n$ according to the principles that the maximum value of space curved face corresponds to the maximum grey level and minimum value of space curved face corresponds to the minimum grey level.

4.3. Feature extraction of grey-scale images

We extract box dimension and information dimension as the features of the grey-scale images. Fractal dimension is a measure for describing the complexity of fractal sets. Box dimension is one of the most commonly used methods of fractal dimension, and it is defined as follows.

\[ D_b = \lim_{r \to 0} \left[ \frac{\ln N_r(F)}{\ln(r^{d})} \right] \quad (9) \]

In Equation (9), $F$ is a nonempty bounded subset belonging to $d$-dimensional Euclidean space $\mathbb{R}^d$. $N_r(F)$ is the minimum number of sets which can contain subset $F$.

Information dimension is defined as Equation (10). $\{X_i\}_{i=1}^N$ is the number of points in nonempty bounded subset $F$ belonging to $d$-dimensional Euclidean space $\mathbb{R}^d$. The value of $N$ is great. $M(r)$ is the number of $d$-dimensional cubes which can contain points $\{X_i\}_{i=1}^N$. $N_i$ is the number of points contained in the $i$th cube, $P_i = N_i/N$.

\[ D_i = -\lim_{r \to 0} \frac{\ln \left( \sum_{i=1}^{M(r)} p_i \ln(p_i) \right)}{\ln(r)} \quad (10) \]

According to Equation (9) and Equation (10), the box dimension and information dimension of grey-scale images are obtained by the least square method and each defects model obtains 28 groups
data. Because box dimension and information dimension are on the different order of magnitude, they need to be standardized by Equation (11). Sample data of box dimension information dimension are shown in Figure 5.

\[ x^* = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \]  

(11)

![Figure 5. Box Dimension-Information Dimension.](image)

4.4. PD pattern recognition by AP-MD

Cluster the sample data of box dimension and information dimension by AP-MD, AP, \( k \)-means and FCM. The results are shown in Figure 6. The most obvious difference of Figure 6 (a), Figure 6 (b). Figure 6 (c) and Figure 6 (d) is that point \( x_o \) is assigned to the category \( P_i \) in Figure 6 (a). Clustering results of AP, \( k \)-means and FCM are very similar. The recognition rates of AP-MD for corona discharge, surface discharge and cavity discharge are 100%, 50%, 65.9%. The recognition rates of other algorithms are 96.4%, 32.1%, 60.8%.

![Figure 6. Clustering results of AP-MD, AP, \( k \)-means and FCM.](image)

5. Conclusion

The \( k \) value has effects on clustering result of AP-MD and the number of iterations. When \( k \leq 4 \) or \( k \geq 27 \), the accuracy of clustering is reduced. The optimal \( k \) value is 10% ~ 15% of the amount of data contained in one category and doesn’t be less than 4.

The clustering result of AP-MD is better than results of AP, \( k \)-means and FCM for data which are distributed regularly. The clustering results of AP, \( k \)-means and FCM for those data are wrong. But it is very applicable to clustering these data.

The recognition rates of AP-MD for corona discharge, surface discharge and cavity discharge are 100%, 50%, 65.9%. Compared with AP, \( k \)-means and FCM, AP-MD improves the accuracy of partial discharge pattern recognition.

References

[1] Wang M H 2005 Partial discharge pattern recognition of current transformers using an ENN IEEE Transactions on Power Delivery 20 (3) pp 1984–90.

[2] Wang K, Yang L J, Liao R J, et al 2011 Partial discharge pattern recognition based on discrete hidden Markov models Transactions of China Electrotechnical Society 26 (8) pp 205–12.

[3] Liu S B, Lv C, Yu J L, et al 2008 Application of Hilbert-Huang transform in pattern recognition for partial discharge of transformers Proceedings of the CSEE 28 (31) pp 114–9.
[4] Rodrigo Mor A, Castro Heredia L C and Muñoz F A 2017 New clustering techniques based on current peak value, charge and energy calculations for separation of partial discharge sources *IEEE Transactions on Dielectrics and Electrical Insulation* **24** (1) pp 340–8.

[5] Tan K X and Li F Q 1996 Some types of artificial neural networks used for recognition of partial discharge *High Voltage Engineering* **22** (4) pp 3–7.

[6] Sun C X, Xu G F, Tang J, *et al* 2005 PD pattern recognition method using box dimension and information dimension as discriminating features in GIS *Proceedings of the CSEE* **25** (3) pp 100–4.

[7] Karivda A 1995 *Recognition of discharges, Discrimination and classification* (Delf: Delft University Press).

[8] Chen W G, Ling Y, *et al* 2012 Method to identify developing stages of air-gap discharge in oil-paper insulation based on cluster-wavelet neural network *Power System Technology* **36** (7) pp 126–32.

[9] Frey B J and Dueck D 2007 Clustering by passing messages between data points *Science* **315** (5814) pp 972–6.

[10] Sun L L, Guo C H, Liu C R, *et al* 2017 Fast affinity propagation clustering based on incomplete similarity matrix *Knowl Inf Syst* **51** (3) pp 941–63.

[11] Wang R P, Shen S G, Chen X L, *et al* 2012 Manifold-manifold distance and its application to face recognition with image sets *IEEE Transactions on Image Processing* **21** (10) pp 4466–79.

[12] Hussein R, Shanban K B and EI-Hag A H 2017 Robust feature extraction and classification of acoustic partial discharge signals corrupted with noise *IEEE Transactions on Instrumentation and measurement* **66** (3) pp 405–13.