USE OF HIDDEN MARKOV MODEL FOR PARTIAL DISCHARGE-LED FAILURE MODELLING

D. Zhou[1,2] and C.R. Li[1] and Z.D.Wang[2*]
[1]North China Electric Power University, Beijing, P.R. China
[2]The University of Manchester, Manchester, UK
*Email: <zhongdong.wang@manchester.ac.uk>

Abstract: This paper presents a hidden Markov model (HMM) based approach to partial discharge-led failure modelling. The approach is based on a continuous probability density, left to right one way hidden Markov model which closely matches the initiation to breakdown transition sequence and development process of discharge-led failure. Initial estimates of the HMM parameters are discussed. Prior knowledge is incorporated to help the training. Results show that the HMM-based model is an effective approach for state recognition and prediction of PD-led failure.

1 INTRODUCTION

Partial discharge (PD) monitoring is of vital importance to indicate the insulation integrity of high voltage equipment, especially for equipment with a complex insulation system such as power transformers. In power transformers, PD can lead to surface tracking on solid insulation materials, which causes irreversible damage and may eventually lead to a breakdown of the oil system in such a way that the insulation properties of the oil can no longer be guaranteed.

With the development of advanced sensors and data-acquisition systems, on-line monitoring of partial discharge becomes possible. Several international standards and guidance have been published on the topics of conventional [1-2] or unconventional [3] partial discharge measurements. However, there is still a lack of effective data analysis and modelling methods for the correct interpretation of the obtained PD monitoring data, especially to predict and prevent PD-led failures.

The realization of auto-recognition of the severity or state of PD process is of great practical value, any change of the state in discharge degradation will be self-announced before the final failure. Thus, early-warning of PD-led failure can be achieved.

Hidden Markov Model (HMM), among various intelligent techniques, shows the most promise in partial discharge progress modelling. Firstly popularized in speech recognition field since 1960s, its usage has been extended to a wide range of applications nowadays, such as partial discharge pattern classification/recognition [4-5] and transformer failure rate prediction [6], etc.

HMM comprises an underlying stochastic process that is not directly observable but can be visualized through another set of stochastic process that produces a sequence of observations. This is in accordance with the nature of partial discharge developing process as that the real underlying developing state cannot be directly recognized without careful analysis of collected discharge signals.

The paper presents an approach to build a partial discharge development model based on HMM. The principle of a HMM and the proposed modelling process for PD developing state recognition and prediction is introduced. The implementation process of the proposed model based on laboratory collected data is illustrated. The modelling results are also presented.

2 GENERAL METHODOLOGY

The hidden Markov model (HMM) is extended from the concept of Markov chain. It is developed to include cases where the observation is a probabilistic function of the state. According to their probability density descriptions, HMMs are further divided into continuous and discrete models, referred as CHMM and DHMM respectively.

Continuous HMM is selected for this research due to the nature of our observation vectors as they are continuously distributed.

Model structure has firstly to be selected. The left-right CHMM is chosen to model the transition between PD's developing states. The justification for using the left-right HMM for PD modelling and prediction is clear. Based on our previous experimental studies [7-8], it is shown that the development of partial discharges in some oil-paper insulated PD models will going through several degradation states before the final insulation breakdown. These transitions are normally ordered, starting from the initiation, and then gradually progressing through successively more severe states before the final breakdown (failure).

The topology of the proposed left-right HMM is shown in figure 1. In which underlying states S_1 , S_2 , S_3 and S_4 are borne with their corresponding physical meanings which are specified as initiation state, developing state I, developing state II and pre-breakdown state respectively.

The CHMM under consideration consists of :

- A finite set of N states $S = \{S_1, \cdots, S_N\}$.

- A state transition probability distribution matrix $A = \{a_{ij}\}$, where a_{ij} is the probability of making a transition to the state S_j from state S_i during the next event, denoted as $a_{ij} = P\{q(t+1) = S_i \mid q(t) = S_j\}$ (*1≤ i, j ≤N, 1 ≤ t < T*), $q(t)$ is the state at time t .

- An observation symbol probability distribution in state S_j , $B = \{b_j\}$,i.e. $b_j(O(t)) = P\{O(t) \mid q(t) = S_i\}$ (*1≤ i ≤ N,1 ≤ t <T*). $O(t)$ is the observation vector at time t. Generally, the probability density function (p.d.f.) is considered to be of a mixture of Gaussian distribution, with M mixture centers.

- An initial state probability distribution $\Pi = \{\pi_i\}$, where $\pi_i = P\{q(1) = S_i\}$ (*1≤ i ≤ N*).

Once N and M together with the three probability measures A , B and Π are specified, CHMM is completely specified. For convenience, the compact notation of $\lambda = (A, B, \Pi)$ is used to denote the complete parameter set of the model.

In PD development modelling, features extracted from the measured signals are observation symbols. The development stages are chosen as the underlying states.

Figure 2 shows the general procedure of using HMM in PD modelling. The whole process mainly consists of two phases: (1) training and (2) recognition and prediction. In either phase, observation data has to be carefully collected; suitable features shall be selected as to best represent the information hidden in the raw data. Details as for this step will be discussed in Section 3.

2.1 Training

The training stage attempts to optimize the model parameters (A, B, Π) to best describe the observation sequences, which is to optimize $P(O \mid \lambda)$, the probability of observation sequence O given model λ .

Baum-Welch re-estimation algorithm, also known as expectation maximization (EM) approach is used here for the training procedure. Details of the mathematical description of this algorithm can be referred to [9-10].

It is worth noting that, since the algorithm is conducted in an iterative procedure. Initial values of HMM parameters shall be carefully selected. Since it is a left-right HMM, the initial probability distribution can be naturally chosen as $\Pi_1 = [1\,0\,0\,0]$. The initial value of the transition matrix is estimated with the method proposed in [10].

What most problematic here is the initial estimation of B . Prior knowledge of the PD development process shall be incorporated in the estimation. Here, reasonable assumptions based on the prior knowledge are made that for each observation sequence, the four underlying states are uniformly distributed. Therefore, the mean and covariance of the observation can be calculated through k-means clustering process. For cases where $M > 1$, the observations are specified as multimodal Gaussian distribution, while in the case of $M = 1$, it is retrieved to unimodal Gaussian distribution.

Suppose that M can be chosen from a series of

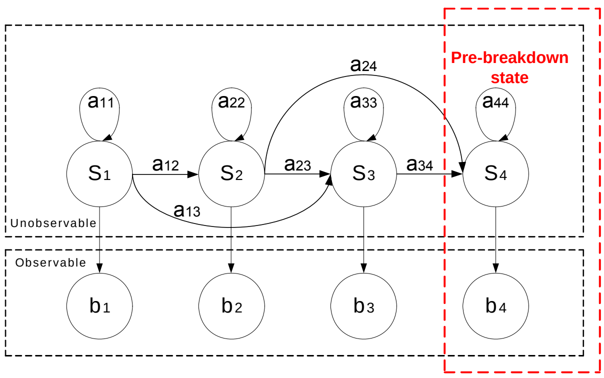


Figure 1: Topology of the proposed HMM

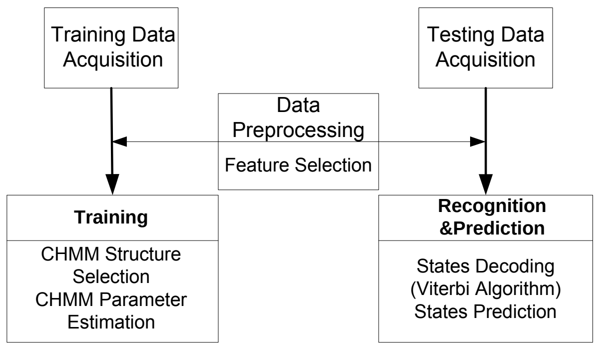


Figure 2: General procedure of using HMM in PD modelling

numbers $M = 1, 2, \cdots, K$. The optimal number, M^* is chosen as the number which maximizes $P(O \mid \lambda)$ to the largest extent, which can be expressed as: $M^* = \arg\max_{1 \le m \le K} P(O \mid \lambda(m))$, where $\lambda(m)$ is the HMM with m mixture centers for its observation probability.

In real application, $\log P(O \mid \lambda)$ instead of $P(O \mid \lambda)$ is chosen to be maximized for its convenience in calculation.

Several experimental trials are conducted in order to choose the optimal value of M. The results are discussed in Section 4.

2.2 Recognition

The recognition stage involves decoding of the underlying states for new observation sequences.

This means to find the optimal state sequence $Q = (q(1), q(2), q(3), \cdots, q(T-1), q(T))$ associated with the given observation sequence $O = (O(1), O(2, O(3), \cdots, O(T-1), O(T))$ and the trained model λ specified in the training stage.

The criteria of optimality here is to search for a single best state sequence through Viterbi Algorithm [9-10].

In the proposed modelling procedure, recognitions are made whenever a new observation is collected, with the features of discharging signals as observations. The underlying PD developing state is then decoded at each observation.

2.3 Prediction

In the prediction stage, the attention is focused on calculating the likelihood of entering the pre-breakdown state in the next observation or next time step given current and previous observations based on specified model λ .

The prediction is made based on the probabilistic structure of the HMM. The probability of a transition to state S_j during the next observation is given by:

$$P[q(t+1) = S_j \mid \lambda] = \sum_{j=1}^{j} P[q(t) = S_i \mid \lambda] a_{ij} = \sum_{i=1}^{j} \alpha_t^*(i) a_{ij} \quad (1)$$

Where: $\alpha_i^*(t)$ is the normalized forward probability at time t for each state S_i . It is calculated from the forward probability $\alpha_i(t)$ with the help of Bayes' rule.

The forward variable $\alpha_i(t)$ is defined as:

$$\alpha_t(i) = P[O(1), O(2), \cdots, O(t), q(t) = S_i \mid \lambda] \quad (2)$$

Indicating the joint probability of the partial observation sequence (until time t), $O(1), O(2), \cdots, O(t)$, and state S_i at time t given the model λ .

The normalized forward probability $\alpha_i^*(t)$, which is the posterior probability, is calculated as:

$$P[q(t) = S_i \mid O(1), \cdots, O(t), \lambda] = \frac{P[q(t) = S_i, O(1), \cdots, O(t) \mid \lambda]}{P[O(1), \cdots, O(t) \mid \lambda]} = \frac{\alpha_t(i)}{\sum_{j=1}^{N} \alpha_t(j)} \quad (3)$$

Using equation (1), the probability of transitioning to each state, including the pre-breakdown state can be computed at each time t. Therefore, the prediction of PD-led failure is made with the calculation result of $P[q(t+1) = S_4 \mid \lambda]$ at each observation time.

3 IMPLEMENTATION

3.1 Data acquisition

Experiments were carried out to obtain sufficient data to train and test the proposed HMM. In our experiments, UHF signals were captured from a transformer PD model as shown in figure 3.

Sequences of UHF signals covering all the evolution states were recorded from the beginning of the PD until the insulation breakdown of the PD model under a pre-specified applied voltage. Altogether, three sets of discharging data with different lengths of discharging time were collected.

3.2 Data preparation and HMM training

To prepare the PD data for the training of the HMM, pulse repetition rate, maximum pulse amplitude and average pulse amplitude of those collected UHF signal sequences were selected as

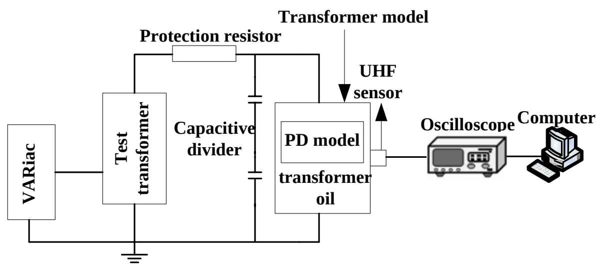


Figure 3: Schematic of experimental setup

the observation features and were calculated in every 30minutes, results are shown in figure 4.

Therefore, the original UHF signal sequences were transferred into the three new observations, with time unit set as 30min. The numbers of observations for sequence1, sequence2, sequence3 are 22, 17 and 14 respectively.

Later on, HMMs are trained with these observations following the procedure showed in figure2.

4 RESULTS AND DISCUSSIONS

4.1 Choices of M value

The number M is initially chosen from 1 to 5, their corresponding $\log P(O|\lambda)$ for each observation sequence are calculated and compared.

Due to the limitation on the amount of training data, both the (1) all data training method, where all three observation sequences are taken as training data; and (2) the leave-one-out method, where the model is trained with one set of the observation sequence being left out of the training sets; are utilized here.

Therefore, for each number of M, with 3 sequences by 4 models, 12 sets of $\log P(O|\lambda)$ are calculated. Altogether, for the 5 M values, 60 results are calculated.

The mean value of $\log P(O|\lambda)$ at each M value is shown in table 1. The result shows that M=3 shall be the optimal choice for our present model since it generates the highest mean value of $\log P(O|\lambda)$.

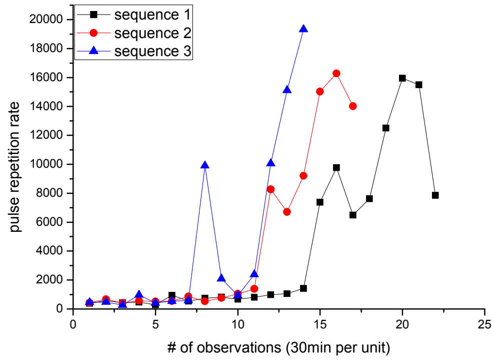
Table 1: mean value of $\log P(O|\lambda)$ v.s. M value

M value	Mean value of $\log P(O\|\lambda)$
M=1	-109.98572
M=2	-105.14081
M=3	**-98.74411**
M=4	-105.92714
M=5	-100.41395

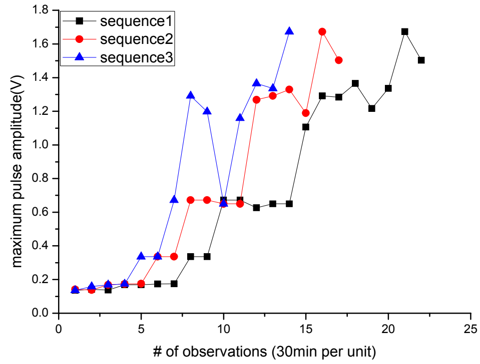
4.2 Recognition and prediction result

A HMM model with sequence 1 and sequence 3 as training data, and sequence 2 as testing data is trained to show the effectiveness of the proposed approach for PD-led failure modelling.

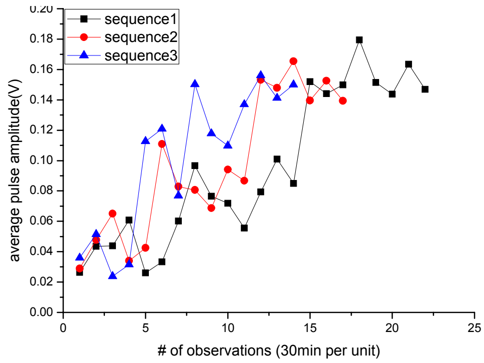
State recognition results of the two training sequences are shown in figure 5. Figure 6 shows the state recognition result versus the number of observations for the training data, sequence 2.



(a). Pulse repetition rate



(b). maximum pulse amplitude



(c). average pulse amplitude

Figure 4: Calculated observation sequences for PD modelling

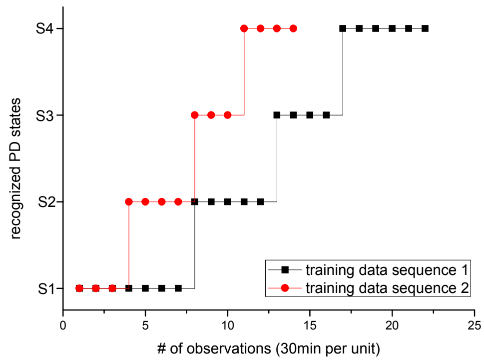


Figure 5: State recognition results of training data

It is showed that the four underlying PD developing state can be successfully recognized. The state transition sequence is similar to the sequences of training data as shown in figure 5.

The time of state transition can be identified, with the 13th observation being identified as the time of transitioning to pre-breakdown state from the previous state. It is 5 observations ahead, or 150 minutes before the occurrence of the final breakdown.

With the pre-failure state being identified, warnings can be given to the system. Therefore, appropriate actions can be taken in time to prevent the PD-led failure events; losses caused by the failure can be prevented.

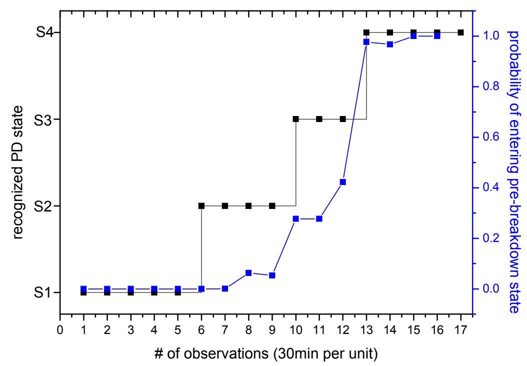


Figure 6: State recognition and prediction result of testing data

Figure 6 also plots the probability of entering the pre-breakdown state during the next observation versus the present observation for sequence 2. It is showed that the predicted probability can effectively track the PD-led failure process with the probability of entering pre-breakdown state increases with the number of observations.

A high probability of $\boldsymbol{P}[\boldsymbol{q}(\boldsymbol{t}+1)=\boldsymbol{S}_4 \mid \lambda]=0.4228$ is predicted at the 12th observation, where $\boldsymbol{t}=12$. The result can warn the system that there is a high probability of failure during the next observation, which is the 13th observation. This matches the recognition result that the sequence enters the pre-breakdown stage at the 13th observation, which in a way proves the effectiveness of the methodology.

With the help of prediction, system warnings can be made one observation time forward. Therefore, reaction time can be furtherly saved for failure prevention actions.

5 CONCLUSION

In this paper, a HMM based modelling procedure is proposed to model the PD-led failure process. The model is capable of making state identification/recognition and predictions for PD-led failures. Experimental study is conducted to show the effectiveness of the proposed modelling procedure. With the help of the proposed model, early-warning of PD-led failures becomes possible. Reaction time can therefore be greatly saved to make time for failure prevention procedures to be taken.

6 ACKNOWLEDGMENTS

We thank the China Ministry of Science and Technology for supporting of this research under National Key Basic Research Program (973 program) with contract number 2009CB724508.

7 REFERENCES

[1] “IEC 60270: High-Voltage Test Techniques-Partial Discharge Measurements”, International Standard, Third Edition, 2000-12.

[2] "IEEE Guide for Partial Discharge Measurement in Liquid-Filled Power Transformers and Shunt Reactors", IEEE Std C57.113-1991, 1992.

[3] "Guidelines for Unconventional Partial Discharge Measurements", CIGRE-WG D1.33, CIGRE Tech. Brochure no.444, 2010.

[4] L. Satish and B. I. Gururaj, “Use of Hidden Markov Models for partial discharge pattern classification”, IEEE Trans. Dielectr. Electr. Insul., Vol. 9, pp. 272-233,2002.

[5] T. K. Abdel-Galil, Y. G. Hegazy, M. M. A. Salama and R. Bartinikas, “Partial discharge pulse pattern recognition using hidden markov models”, IEEE Trans. Dielectr. Electr. Insul., Vol. 11, pp. 715-723, 2004.

[6] Pathak, J., Yong Jiang, Honavar, V., McCalley, J.,” Condition data aggregation with application to failure rate calculation of power transformers”, System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on, 04-07 Jan. 2006, pp. 241a-241a.

[7] Wang Hui, Li Cheng-rong, Sheng Kang, Miao Yinyin, "Experimental study on the evolution of surface discharge for oil-paper insulation in transformers", Electrical Insulation and Dielectric Phenomena, 2009. CEIDP'09. IEEE Conference on, 18-21 Oct.2009, pp. 405-408.

[8] Yuan Yang, Li Chengrong, Zheng Zhong, Liu Shaoyu, Ma Jixian, "Lab investigation of PD development in transformer winding", Electrical Insulation (ISEI), Conference Record of the 2010 IEEE International Symposium on, 6-9 June 2010.

[9] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech

recognition", Proceedings of the IEEE, Vol.77, pp.257-286, 1989.

[10] Waleed H. Abdulla and Nikola K. Kasabov, "The concept of hidden markov model in speech recognition", The Information Science Discussion Paper Series, University of Otago, May 1999, http://hdl.handle.net/10523/997, last accessed: May 11, 2011