APPLICATION OF NEURAL NETS FOR MODELLING PARTIAL DISCHARGE PHENOMENON

Saradindu Ghosh
Department of Electrical Engineering
Regional Engineering College
Rourkela, India

Kishore N K
Department of Electrical Engineering
Indian Institute of Technology
Kharagpur, India.

Abstract **- Classical approach to modelling is quasi-empirical relationship based on experiments on single artificial voids of well defined geometry. Such methods restrict the validity to the range of inputs considered. Keeping all this in view, this work attempts at applying Artificial Neural Network (ANN) for the modelling in order to exploit flexibility of ANN modelling with a short time for development and reasonably high accuracy. The results indicate good agreement of the estimates with the published values with a MAE of as low as 1 %.**

I. Introduction

The continuity of electricity is of utmost importance to all our activities in day to day life. This is ensured by having reliable insulation system in the power apparatus. Partial Discharge (PD) testing is gaining wide acceptance as a non-destructive test tool for assessment of insulation [1], [2]. By nature any discharge or breakdown taking place in the bulk of the insulation without bridging the electrodes is termed as 'Partial Discharge'. Voids or gaseous inclusions in the bulk of the insulation, which are distributed randomly either due to natural defects or by way of manufacturing are the potential sites of such discharges. Hence, any attempt at modelling the phenomenon in a void would go a long way in assessing the insulation quality.

This paper details PD modelling from two angles :

1. **from fundamental view point, considering the modelling for PD inception voltage (PDIV) and PD extinction voltage (PDEV) as a function of void dimensions for single artificial disk shaped void in sheet samples and**
2. modelling of PDIV for a power apparatus, viz., Epoxy-resin Post Insulators.

Artificial Neural Network (ANN) with adaptivity and nonlinearity are well suited to function estimation tasks, particularly where the equation describing the function is unknown. In function estimation application ANN acts as a model which stands for the system it represents, typically to predict or control it. Among the various ANNs presented so far, the present work employs a Multilayer Feedforward Network (MFN) with Back Propagation Algorithm (BPA) with and without adaptive learning.

The effect of learning rate, η and momentum constant, α on the convergence property of the learning process is extensively studied and the best combination is identified. In addition, number of nodes in the hidden layer are also varied to see their effect on the convergence rate. Further, an attempt is made to assess the effect of number of hidden layers on the convergence characteristics.

II. Multilayer Feedforward Network

Artificial Neural Networks are characterized by their topology, that is, by the number of interconnections, the node characteristics that are classified by the type of nonlinear elements used and kind of learning rules employed [3], [4], [5]. These rules specify an internal set of weights and indicate how the weights should be adapted during use to improve the performance.

The ANN is composed of an organised topology of Processing Elements (PE), called neurons. Although a single neuron can perform certain simple functions, like, multi-input nonlinear device, the power of neu-

0-7803-5812-0/00/$10.00©2000IEEE

ral computation comes from connecting neurons into networks. In MFN, the PEs are arranged in layers and only PEs in adjacent layers are connected as shown in the Figure 1.

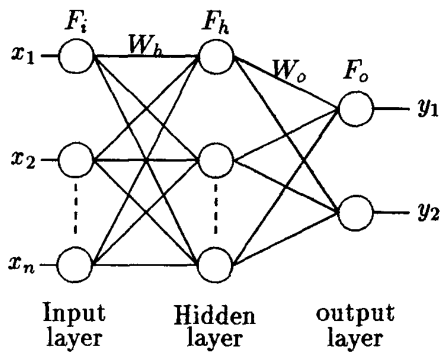

Fig. 1. Schematic diagram of a Multilayer Feedforward Network

MFN is the most popular connectionist model that has been playing a central role in the application of neural networks. It consists of an input layer, one or more hidden layers and an output layer. The information propagates only in the forward direction and there are no feedback loops. The number of PEs and hidden layers may be chosen carefully, so as to optimize the performance of the network. Each connected pair of neurons are associated with an adjustable value that is referred to as the weight or synapses. The total input to a neuron is the weighted sum of neuron outputs from the previous layer. A neuron's output is computed by feeding its threshold input through a nonlinear squashing function to limit its output. A bias term is often incorporated at each neuron to improve the convergence.

In order to obtain bounded output from PEs a sigmoidal activation function is chosen such that the output is limited to (0, 1) for an input range of ($-\infty, \infty$).

A. Learning Process in ANN

Essentially, the ANN evolves a nonlinear mapping between the input and output patterns during the learning phase. The learning process in this work, employs a learning rule called the BPA, alternately known as Generalised Delta rule, which is an iterative gradient technique that performs the input-output mapping by minimizing the cost-function. Most learning procedures use a gradient descent method to find an optimal set of connection weights. The training examples are presented over and again in random order and the weights and biases are updated depending on the negative gradient of the respective output error until the error reaches an acceptable minimum. This is called a supervised learning scheme.

B. Scaling of Input-output Pattern

Scaling of input-output patterns is an important aspect of the design of the ANN. A suitable range of scaled input patterns often results in rapid convergence and good recognition ability, which may otherwise lead to saturation. Also, it is obvious that for the sigmoidal activation function the output range of the ANN must lie within (0,1). Thus, the input-output pattern are scaled in the range of (0,1) before the initiation of the training of the ANN.

In this case each input or output data x_i is normalized as p_i before being fed to the ANN according to

$$p_i = \frac{x_i}{x_{max}}, \qquad i = 1, 2, \cdots, n \tag{1}$$

where x_i and x_{max} are the actual data and the maximum value of the input or output patterns respectively

C. On-line and Batch Learning

There are two basic approaches adopted to minimize the global error function, E [3]. They are :

1. On-line learning and
2. Batch learning.

In an on-line learning, the patterns are presented sequentially, usually in a random order. For each learning example the weights are changed proportional to the respective negative gradient of the local error function, E_p.

Now, if the learning rate, η is sufficiently small, on-line procedure minimizes the global error function, $E = \sum_p E_p$ [16].

On the other hand, in a batch learning, the total error function, E is minimized in such a way that the weight changes are accumulated over all learning examples before the weights are actually changed.

The relative effectiveness of the on-line and batch learning procedures is highly dependent on the problem, but the on-line algorithm seems superior in most cases primarily because [3] :

1. On-line procedure introduces some randomness (noise) that often helps in escaping from a local minima.
2. On-line procedure is faster and more effective than the standard batch procedure.

D. Evaluation criterion

There are two qualitative measurements revealing the status of the learning process or the status of function estimation [5], [6]. One is the mean sum squared error of the training data (E_{tr}), the other is the mean sum squared error of the test data (E_{ts}). They are defined as

$$E_{tr} = \frac{1}{p}\sum_{p}(T_p - O_p)^2 \qquad (2)$$

where T_p and O_p are the target and the calculated output respectively corresponding to the training data pattern and p is the number of training patterns, and

$$E_{ts} = \frac{1}{s}\sum_{s}(T_s - O_s)^2 \qquad (3)$$

where T_s and O_s are the target and the calculated output respectively computed by forward propagation of the corresponding test data pattern and s is the number of test patterns. The network tends to interpolate training data as E_{tr} approaches zero. The E_{tr} tells how well the network is adapted to fit the training data only, even if the data are contaminated.

On the other hand, the E_{ts} indicates how well a trained network behaves on a new set of data, which are not included in the training set. This behaviour is known as generalization. Since, it is assumed that there are no errors in the test data, E_{ts} correctly reflects, how well a trained network has learned from the training data to approximate the underlying function. Therefore, E_{ts} is the criterion for an evaluation of the performance of a trained network in function estimation.

As a matter of fact, in this study, some models are trained on the basis of E_{tr}, while others are on the basis of minimum E_{ts}. As E_{tr} reduces with number of iterations, the models based on E_{tr} are trained only up to a certain number of iterations.

E. Adaptive Backpropagation Algorithm

It is very clear that η and α have a very significant effect on the learning speed of the BPA. A large value of η results in a faster convergence but often leads to oscillations. Whereas, a small value of the η stabilizes the process but results in a slower convergence and increases the ANN susceptibility of getting entrapped in a local minima. Similarly, an increase in α when connection weights are updated in correct directions (that is, if there is a reduction in the error) improves convergence. On the other hand, if the update direction is wrong (that is, if there is an increase in the error), the value of α should be reduced to improve the convergence.

Research into dynamic change of the η and α of the BPA has been carried out by many authors [7], [8]. A simple method of updating the η and α is presented by Jacob [7]. Yu, Chen and Cheng [8] observe that the optimal η varies almost randomly from iteration to iteration.

Based on this approach and with a knowledge of the immediate preceding error signal, an adaptive BPA with a dynamic η and α developed for the present work can be given as follows:

$$\begin{aligned} \eta(k+1) &= 0.99\eta(k); \ \text{if}\Delta E(k) < 0 && (4) \\ &= \eta(k); \ \text{if}\Delta E(k) \geq 0 && (5) \\ \text{and } \alpha(k+1) &= 1.005\alpha(k); \ \text{if}\Delta E(k) < 0 && \\ &= 0.999\alpha(k); \ \text{if}\Delta E(k) \geq 0 && (6) \end{aligned}$$

where

$$\Delta E(k) = E(k) - E(k-1) \qquad (7)$$

and $E(k)$ is the training error at k th iteration.

The algorithm begins with an initial value of η and gradually reduces as the learning progresses. Starting value of α is chosen as 0.9. The maximum and minimum value of α is specified as 0.99 and 0.5 respectively.

III. Results and Discussions

In all four models are presented here which correlate the PD quantities, namely, PDIV , PDEV and PD inception stress to the PD parameters, such as, insulation thickness, void diameter, void depth, pressure of the gas inside the void and the physical dimensions of the post insulators. The effect of void location and electrode material on PD inception voltage are also studied.

In Model - I, PDIV across void, v_i' is estimated as a function of PD parameters, like, insulation thickness, t, void depth, t_1 and void diameter, d. The network is trained employing 10 sets of input - output data taken from the literature [9]. Since, only a few sets of input-output patterns are available here, same data sets are used for the testing of the network also, on completion of the training. It is found that modelled values closely follow the calculated value of the PDIV across the void and a Mean Absolute Error (MAE) of 1.3133 % is obtained [10].

In Model - II, modelling of discharge inception voltage across void, v_i' and inception stress E_i' based on void depth and gas pressure are the prime consideration. The requisite training data are obtained from the experimental studies, due to Hall and Russek [11]. The learning process of the network is carried out with 29 input patterns. On completion of the training, the ANN is tested with 6 patterns, which are excluded while training. Detailed studies are carried out to determine the ANN parameters which give the best results. The results show that MAEs of the ANN outputs are found to be 3.26 % and 1.98 % respectively for void inception voltage, v_i' and stress, E_i'. Further, the estimates of void inception voltage by ANN are compared with the analytical results obtained from empirical relations proposed by other researchers [11], [12] and are found to be in good agreement [13].

In Model - III, an attempt is made to estimate the PDIV of solid dielectrics based on void diameter, d, void depth, t_1 and dielectric sample thickness, t. The ANN is trained with 10 sets of input-output patterns based on the data, due to Mason [9], consisting of the calculated values of PDIV, v_i and PD parameters, namely, void diameter, d, void depth, t_1 and dielectric sample thickness, t. These data sets correspond to the void location within the bulk of a dielectric sample (polythene) placed between two brass electrodes. On completion of training, the v_i value is estimated by the network with the data corresponding to a void located between an electrode (brass) and the dielectric sample. Simultaneously, the network is also tested with data corresponding to the other electrode material, that is, when dielectric sample is placed between steel electrodes. The main intention of such a modelling is to observe the effect of void location and the electrode material on PD. An examination of the estimates reveals that the PDIV does not depend on the position of void location and electrode material used but depends on the PD parameters [14], which validates the observations of earlier researcher [11].

In Model - IV, PDIV is estimated as a function of physical dimensions of the post insulator, namely, creepage length and electrode spacing. The modelled results show a very close relationship with the experimentally measured value [15]. The proposed modelling of PDIV, v_i is carried out using the data obtained from the experiments under 50 Hz ac supply. Out of the 17 sets of experimental input - output patterns, 12 sets of input - output patterns (arbitrarily chosen) are utilized to train the network and the remaining 5 sets are used for testing purposes

The combination of ANN parameters for the best results in each of the models have been identified. Table I shows the combination of ANN parameters obtained for the best results for the models considered.

A. Comparison of On-line and Batch Procedure of Training

Figure 2 shows the $MSSE$ distributions of the test data as a function of number of iterations for on-line and batch procedure of training the network for Model- III. From the figure it is clear that the on-line procedure is more effective for the proposed estimation work.

B. Comparison of Adaptive Algorithm with Conventional BPA

To see the effectiveness of the adaptive algorithm over the conventional BPA, the variation of $MSSE$

TABLE I

ANN PARAMETERS OBTAINED FOR THE BEST RESULTS.

Model Number	η	α	Number of Hidden Layers	Number of Neurons	Number of iterations	Evaluation criterion	Algorithm adopted
I	0.25	0.9	1	3	2000	E_{tr}	Batch-mode
II	0.9	0.95	2	6	2500	E_{tr}	Batch-mode
III	0.25	0.9	1	6	3000	E_{ts}	On-line
IV	0.1	0.9	1	6	200	E_{ts}	Adaptive

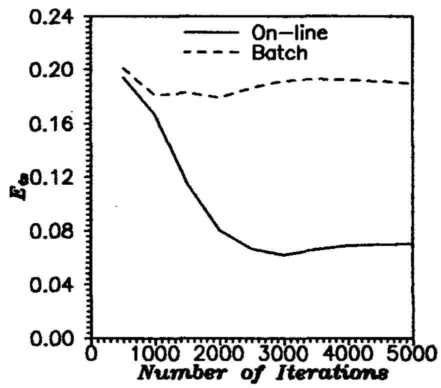

Fig. 2. Variation of Mean sum squared error as a function of iterations for On-line and Batch procedure of learning (Model -III

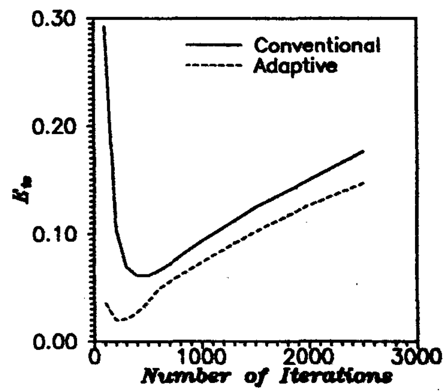

Fig. 3. Variation of $MSSE$ of the test data as a function of the Number of iterations for Conventional and Adaptive BPA

of the test data as a function of the number of iterations for Model- IV is presented in Figure 3. As can be seen, the network, converges much faster in case of adaptive algorithm. The minimum value of E_{ts} is also lower. A minimum value of $E_{ts} = 0.0197$ is obtained after 200 iterations in case of adaptive algorithm as compared to a minimum value of E_{ts} = 0.0610 after 400 iterations in case of conventional BPA.

Finally, the $PDquantities = f(PDparameters)$ for the test data are calculated simply by passing the input data in the forward path of the network and using updated weights of the network. A comparison of modelled and experimental results indicates that ANN can be very well employed for estimation of PD quantities as a function of PD parameters.

IV. Conclusions

1. Though Rumelhart and McClelland [16] suggest learning rate, $\eta = 0.25$ and momentum factor, $\alpha = 0.9$ yield good results for most applications, this work indicates that it is not always so, which is in agreement with the finding of Satish and Zaengl [17] and Chakravorti and Mukherjee [18], that is, best results are seen to yield even for other values of η and α.
2. This study indicates that in MFN, the upper-bound on the number of neurons in the hidden layer follow Hecht-Nielsen [5] criterion.
3. Comparative analysis of the modelled results with the results obtained from the empirical relations given by earlier researchers demonstrate the effectiveness of ANN in modelling an insulation system with unknown nonlinear relation-

ship. On a comparison with the experimental data, the estimates of void inception voltage using ANN are found to be within a MAE of 3.26 % while the MAE of the estimates obtained using the empirical relations proposed by Hall and Russek [11] and Bania and Raghuveer [12] are 4.89 % and 5.77 % respectively.

4. An ANN can model effectively the PD quantities within a small MAE as small as 0.65 % even if the number of input patterns are small.
5. The on-line procedure of learning is more effective for function approximation as compared to batch procedure of learning. This is in agreement with Cichocki and Unbehauen [3].
6. The adaptive learning has definitely a better convergence effect as well as sometimes the accuracy of the modelling data increases. Moreover, one need not worry about the choice of the learning parameters, that is, the learning rate and the momentum constant.

V. ACKNOWLEDGEMENTS

The authors wish to thank the authorities of the Regional Engineering College, Rourkela and the Indian Institute of Technology, Kharagpur for encouragement in this work and permission to present this paper. The authors are grateful to the organisers of ICIT'2000 for inviting them for a contribution.

REFERENCES

[1] R. Bartnikas and E. J. McMahon, Eds., *Engineering Dielectrics, vol. 1, Corona Measurement and Interpretations*, ASTM, STP669, 1979.

[2] F. H. Kreuger, *Discharge Detection in High Voltage Equipment*, Heywood, London, 1964.

[3] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, John - Wiley, Stuttgart, 1993.

[4] R. P. Lippman, "An Introduction to Computing with Neural Nets" , IEEE ASSP Magazine, 1987, pp. 4-22.

[5] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley publishing Company, Inc., USA, 1990.

[6] D. S. Chen and R. C. Jain, " A Robust Back Propagation Learning for Function Approximation " , IEEE Trans. on Neural Networks, vol. 5, no. 3, 1994, pp. 467 - 479.

[7] Jacob, R. A., "Increased Rates of Convergence through Learning rate Adaptation" , Neural Networks, vol. 1, no. 4, 1988, pp. 295-308.

[8] Yu, X. H. and Chen, G. A. and Cheng, S. X., "Dynamic Learning Rate Optimization of the Backpropagation Algorithm" , IEEE Trans. on Neural Networks, vol. 6, no. 3, 1995, pp. 669-677.

[9] J. H. Mason, "The Deterioration and Breakdown of Dielectrics Resulting from Internal Discharges" , Proc. IEE, vol. 98, part I, 1951, pp. 44-59.

[10] Ghosh, S. and Kishore, N. K., "Modelling of Discharge Inception and Extinction Voltages in Dielectric Voids using Artificial Neural Network" , *in 5th Int. Conference on Properties and Applications of Dielectric Material*, Seoul, Korea, 1997.

[11] Hall, H. C. and Russek, R. M., "Discharge inception and extinction in dielectric voids" , Proc. IEE, vol.101, 1954, pp. 47-58.

[12] Bania, J. W. and Raghuveer, M. R., "A new approach to define internal partial discharge intensity" , IEEE Tr. on Electrical Insulation, vol. EI-21, no. 1, 1986, pp. 79-87.

[13] Ghosh, S. and Kishore, N. K., "Application of Artificial Neural Network for Modelling of Discharge Inception Voltage" , *in Int. Conference on Electrical Insulation and Dielectric Phenomena*, Minneapolis, USA, 1997.

[14] Ghosh, S. and Kishore, N. K., "Estimation of Partial Discharge Inception Voltage using Artificial Neural Network" , *in Fourth Workshop and Conference on EHV Technology*, Bangalore, India, 1998.

[15] Ghosh, S. and Kishore, N. K., " Modeling PD Inception Voltage of Epoxy Resin Post Insulators using an Adaptive Neural Network" , IEEE Trans. on Dielectrics and Electrical Insulation, vol. 6, no. 1, 1999, pp. 131-134.

[16] Rumelhart, D. E.and McClelland J. L.,(editors),*Parallel distributed processing : explorations in the microstructure of cognition*, vol.1, MIT press, 1986.

[17] Satish, L and Zaengl, Walter. S., "Artificial Neural Networks for Recognition of 3-d Partial Discharge Patterns" , IEEE Trans. on Dielectrics and Electrical Insulation, vol. 3, no. 2, 1994, pp. 265-275.

[18] Chakravorti, S. and Mukherjee, P. K., "Application of Artificial Neural Networks for Optimization of Electrode Contour" , IEEE Trans. on Dielectrics and Electrical Insulation, vol.1, no. 2, 1994, pp. 254-264.