Chiang Mai J. Sci. 2009; 36(1) : 83-91
www.science.cmu.ac.th/journal-science/josci.html
Contributed Paper

Artificial Neural Networks Parameters Optimization with Design of Experiments: An Application in Ferromagnetic Materials Modeling

Wimalin Laosiritaworn*[a], and Nantakarn Chotchaithanakorn [b]

[a] Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand.

[b] Department of Industrial Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.

*Author for correspondence; e-mail: wimalin@hotmail.com

Received: 9 November 2008
Accepted: 22 December 2008.

ABSTRACT

This paper focused on the application of design of experiments to determine optimize parameters for multilayer-perceptron artificial neural network trained with back-propagation for modeling purpose. Artificial neural networks (ANNs) for modeling have been widely used in various fields because of its ability to 'learn' from examples. The accuracy of ANN model depends very much on the setting of network parameters, such as number of neurons, number of hidden layers and learning rate. Most literatures in this area suggested trial-and-error method for parameters setting which are time consuming and non economical, whereas the optimal setting cannot be guaranteed. Consequently, design of experiment techniques is generally required to optimize various processes. In this paper, as a case study, it was used to find optimum setting of ANN trained to model ferromagnetic material data. Interested characteristic was finite-sized ferromagnetic Curie temperature obtained from Monte Carlo simulation on two dimensional Ising spins. The results indicated that design of experiments is a promising solution to the mentioned problem. The issues arising from this case were also discussed.

Keywords: artificial neural networks, design of experiments, ferromagnetic materials, curie temperature.

1. INTRODUCTION

The applications of artificial neural networks (ANNs) have been reported in literature in various areas for example, manufacturing and business [1 - 5]. This is due to their ability to learn and generalize from examples, their robustness, and fault tolerant. Most literature related to ANNs focused on specific applications and their results rather than the methodology of developing and training the networks. In general, numbers of ANNs' parameters, such as number of hidden node and hidden layer, have to be set during training process and these setting are very crucial to the accuracy of ANNs model. Trial-and-error method is usually used to determine the appropriate setting of these parameters.

Design of experiment (DoE) is a statistical technique widely used to study the relationship between factors affecting the outputs of the process. It can be used to systematically identify the optimum setting of factors to obtain the desired output. In this paper, it was used to find the optimum setting of ANNs' parameters in order to achieve minimum error network. Interested ANNs parameters were number of neuron in hidden layers, momentum, and learning rate. A case study of ferromagnetic materials modeling was used to illustrate the proposed method. The results show that DoE can be used to find better setting of ANNs, which not only results in minimum error, but also significantly reduces training time and effort in the modeling phases.

2. BACKGROUND AND RELEVANT LITERATURES

2.1 Artificial Neural Networks

ANN is by definition "an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection, or *weights*, obtained by a process of adaptation to, or learning from, a set of training patterns" [6]. The ANN can be 'trained' to model relationships between input and output parameters from examples of the known inputs and their corresponding outputs. These input and output are presented to the network using neurons located in input and output layers respectively (e.g. see Figure 1). Neurons are connected where the strength of connections is indicated by "weight". In the training process, a set of examples of input-output pairs are passed through the model and the weights are adjusted in order to minimize error between the answers from the network and the desired outputs. This weight alteration procedure is controlled by the learning algorithm. Learning algorithm adopted in this study is the back-propagation algorithm as it is the most extensively used. In this algorithm, the used parameters are defined as the following. $x_j^{[s]}$ is the current output state of the j^{th} neuron in layer s, $w_{ij}^{[s]}$ is weight on connection joining i^{th} neuron in layer (s-1) to j^{th} neuron in layer s, $I_j^{[s]}$ is the weighted summation of inputs to the j^{th} neuron in layer s, l_{coef} is learning coefficient, and Momentum is the momentum coefficient. Then, the back propagation algorithm is evaluated via the procedures:

1. An input vector i is fed to the input

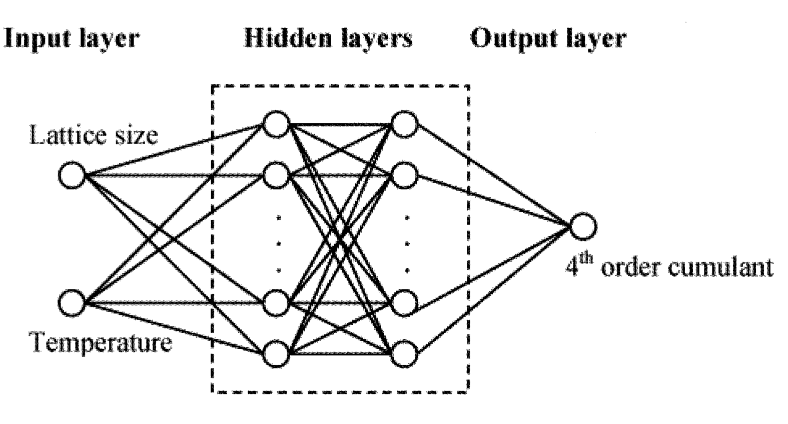


Figure 1. ANN model of ferromagnetic material.

layer of the network through the output layer to obtain an output vector $\underline{o}$. At this stage, the summation of inputs i_j and output x_j of each neuron are also be calculated.

2. For each neuron in the output layer, the scaled local error $e_k^{(o)}$ is calculated from

$$e_k^{(o)} = (d_k - o_k) \cdot f'(I_k), \tag{1}$$

where the delta weight is calculated from

$$\Delta w_{ij}^{[s]} = l_{coef} \cdot e_j^{[s]} \cdot x_i^{[s-1]} + \boldsymbol{momentum} \cdot \Delta w_{ji}^{[s]}. \tag{2}$$

3. For each layer, the scaled local error $e_j^{[s]}$ and delta weight are repeatedly evaluated as

$$e_j^{[s]} = x_j^{[s]} \cdot (1.0 - x_j^{[s]}) \cdot \sum_k (e_k^{[s+1]} \cdot \Delta w_{kj}^{[s+1]}). \tag{3}$$

4. The delta weights are added to the previous weights to update all weights in the network. Then, the whole process is repeated until error is smaller than the stopping criteria.

2.2 Optimization of ANNs Parameters Using DoE

Design of experiment was first developed in the early 1920s by Sir Ronal Fisher to determine the effect of multiple factors on the outcome of agricultural trials. The work of Refs [7, 8] and other statistician with an interest in this area have provided a firm foundation for practitioners. Factorial experimental design which experimental trials are performed at all combinations of factor levels was used in this study.

The applications of DoE techniques to optimize ANNs parameter were reported in literatures [9 - 13]. Factors that often found to have the significant effect on ANNs accuracy are number of neurons in hidden layers, learning rate and momentum as shown in Table 1. Therefore, these factors were carefully chosen to optimize the ANN's modeling in this paper.

Table 1. Research on NN parameters optimization.

Ref.	Applications	Significant factors
[9]	Super plastic forming process modeling	Hint, %of validation, Transfer function
[10]	10 data sets from various sources	Number of hidden neurons, learning rate, momentum term
[11]	Wood veneer inspection	Learning rate, Number of neuron in hidden layer 1
[12]	Vibration suppression	Number of input, number of training pattern
[13]	The selection of manufacturing policies	Input representation, Training sample size

3. THE CASE STUDY : THE FINITE-SIZE CURIE TEMPERATURE FROM THE MONTE CARLO SIMULATION ON TWO DIMENSIONAL ISING SPINS

3.1 Training Data

In this work, a case study of using ANNs to model ferromagnetic materials was studied. The relationships among parameters obtained from Monte Carlo simulation on ferromagnetic Ising spins in two dimensions were investigated. ANN model used in the paper is shown in Figure 1. The training data were obtained from Monte Carlo simulation on two dimensional Ising spins. The input data were system size L (where the total number of spins is $N = L \times L$) and the simulated temperature T. The output data was the fourth order cumulant of the magnetization U_L i.e. $U_L = 1 - \langle m^4 \rangle / \langle m^2 \rangle^2$ [14], where $m = \sum_i S_i / N$ is the average magnetization per spin and $S_i = 1$ is the Ising spin. At the critical temperature, U_L becomes independent of L, so for differing sized L and L', $(U_L / U_{L'})_{T=T_C} = 1$ [14]. Nevertheless, owing to finite size effects, the cumulant curves obtained for different L do not exactly cross at a same temperature. Therefore, the critical temperature is generally estimated from $T_c(b = L/L')$ at the limit $(\ln b)^{-1} \rightarrow 0$ [14]. As a result, to accurately locate the crossing point where $(U_L / U_{L'})_{T=T_C} = 1$ is not trivial. It is required to vary lots of simulated temperatures until the required crossing condition is satisfied. General numerical interpolation techniques, e.g. cubic spline or polynomial interpolations, should be avoided since they are very susceptible to errors if the data is of some uncertainties.

Therefore, it is of interest if there are other techniques which can be used to help finding the cumulant crossing at some good accuracy but do not require much computer resource. As a result, this work considers the use of artificial neural network to interpolate among cumulant data, obtained from the Monte Carlo simulations, to find the crossing point. Each cumulant data point is calculated from small number of spin configurations, i.e. 5000 mcs, with varying system sizes L and temperatures T. In this work, the system sizes were varied from 10 to 50 in steps of 5. The temperature was varied from 2.0 to 2.5 in steps of 10^{-5} J/k_B to cover the analytic two dimensional critical temperature of the Ising model where $T_c \approx 2.269$ J/k_B [15].

Artificial neural networks were firstly trained with 250 patterns of the Monte Carlo data. 75% of these 250 patterns were used for training and the rest 25% was used for testing the network performance to prevent the network from overtraining.

3.2 Experimental Setting

In this paper, the factors that affect the ANNs accuracy, which can be measured by root mean square error (RMSE), were studied. Four factors were included in the experiments as they are factors that often found to be important as reported in literatures. These factors and their setting are shown in Table 2. Two levels of number of neurons in hidden layer 1 were tested which are 2 and 10 neurons. This setting is due to the prior experiments which results shown that more than 10 neurons does not improve much of the network accuracy. Reference [11] suggested that number of neuron in the second layer

Table 2. Factors and their levels.

Factor	Level 1	Level 2
Number of neurons in hidden layer 1 (Factor A)	2	10
Number of neurons in hidden layer 2 (Factor B)	0	6
Learning rate (Factor C)	0.001	0.1
Momentum (Factor D)	0.0	0.8

Table 3. Experimental settings and results.

Run Order	Factor A	Factor B	Factor C	Factor D	RMSE
1	10	0	0.1	0	0.0034
2	10	6	0.001	0	0.0205
3	2	6	0.1	0.8	0.0020
4	2	0	0.001	0	0.0222
5	2	6	0.001	0.8	0.0056
6	2	0	0.1	0	0.0149
7	10	0	0.1	0	0.0030
8	2	6	0.1	0	0.0053
9	2	6	0.001	0	0.0242
10	10	0	0.1	0.8	0.2953
11	10	0	0.001	0.8	0.0085
12	2	0	0.001	0	0.0212
13	2	6	0.001	0.8	0.0066
14	2	0	0.1	0.8	0.0146
15	10	6	0.1	0.8	0.0016
16	2	6	0.001	0	0.0063
17	10	0	0.001	0.8	0.0082
18	2	0	0.1	0	0.0148
19	2	0	0.001	0.8	0.0209
20	10	0	0.1	0.8	0.2953
21	10	0	0.001	0	0.0192
22	2	0	0.1	0.8	0.0146
23	2	0	0.001	0.8	0.0188
24	2	6	0.1	0	0.0082
25	10	6	0.1	0	0.0029
26	10	6	0.001	0.8	0.0045
27	10	0	0.001	0	0.0224
28	10	6	0.1	0	0.0021
29	2	6	0.1	0.8	0.0041
30	10	6	0.001	0.8	0.0044
31	10	6	0.1	0.8	0.0021
32	10	6	0.001	0	0.0057

should start from 0. This factor was then set at 0 neuron for level 1 and 6 neurons for level 2. Learning rate was set at 0.001 and 0.1, while momentum was set at 0.0 and 0.8.

Experimental setting was shown in Table 3. Full factorial design for four factors, namely 2^4 designs were used, results in 16 experimental runs. The experiments were run at two replicates per each setting. As a result, a total of 32 runs were conducted. Experiments were conducted according to run order. For example, the first experiment was carried out at 10 neurons in the first hidden layer, 0 neuron in the second layer, 0.1 learning rate and momentum term equals to 0. All 32 experiments were performed according to the setting and experimental results (RMSE) were shown in the last column in Table 3.

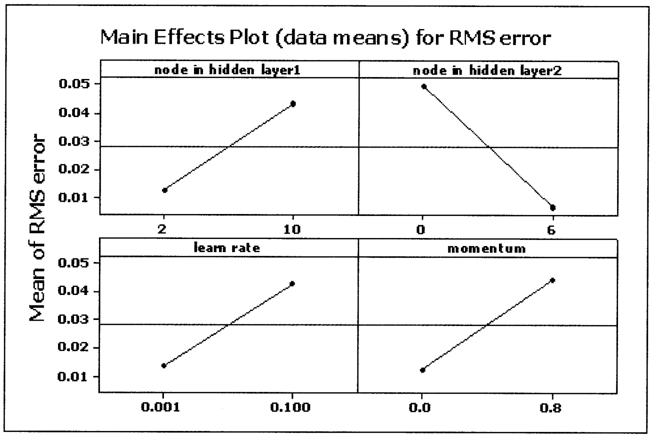
3.3 Experimental Results and Discussion

Analysis of variance (ANOVA) was carried out based on the unseen testing data at 95% confidence level. The analysis was prepared by MINITAB software. Table 4 shows the estimated effects of each parameters and coefficient for yield. Parameters uses in this table are the parameters defined in table 2. The term A*B is the interaction between factors A and B In this analysis, only two-way interaction was considered.

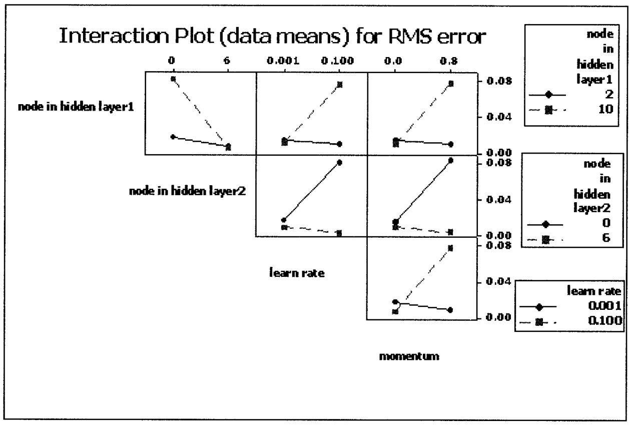
Each column in table 4 contains information regarding the determination of the significant of each term to the response (RMSE). The first column is the name of term analyzed. The second column (Effect) refers to the estimated effects for each factor. Effect was calculated from Coef×2. 'Coef' refers to the estimation of the population regression coefficients in a regression equation. 'SE Coef' is the standard error which refers to the standard deviation of the estimate of a regression coefficient. The smaller value indicates a more precise estimate. As shown in table 4, the 'SE Coef' of all terms is small, which suggest that the data can precisely estimate the coeffcient's unknown values. 'T' is the statistical t-value while 'P' is the statistical p-value. The p-value was used to determine which of the effects are significant. In this study 95% confidence was used therefore terms that have p-value lower than 0.05 are significant. Table 4 showed that all factors and

Table 4. Estimated effects and coefficients for yield.

Term	Effect	Coef	SE Coef	T	P
Constant		0.02825	0.000745	37.92	0.00
A	0.03093	0.01547	0.000745	20.77	0.00
B	-0.0432	-0.0216	0.000745	-29.00	0.00
C	0.02906	0.01453	0.000745	19.51	0.00
D	0.03192	0.01596	0.000745	21.43	0.00
A*B	-0.03323	-0.01662	0.000745	-22.31	0.00
A*C	0.03499	0.01749	0.000745	23.49	0.00
A*D	0.03565	0.01783	0.000745	23.93	0.00
B*C	-0.03527	-0.01764	0.000745	-23.68	0.00
B*D	-0.03746	-0.01873	0.000745	-25.15	0.00
C*D	0.03995	0.01998	0.000745	26.82	0.00



a) Main effect plot of RMSE



b) Interaction plot of RMSE

Figure 2. Main effect and interaction plot.

their interactions are significant. The result is consistent with the result reported in [10]. The effect of these factors are shown in Figure 2.

Figure 2a) shows the main effect plots of RMSE. The figure indicates that number of neurons in hidden layer 1 at low level (2 neurons) results in lower RMSE. On the other hand, number of neurons in hidden layer 2 at high level (6 neurons) has a better result. Learning rate and momentum terms at lower level (0.001 and 0.0 respectively) result in best accuracy. However, as indicated in Table 4, all interactions among parameters are significant. Therefore, the best setting can not be determined from the main effect plot and the interaction plot (Figure 2b) has to be taken into account.

Figure 2b) shows the two-factor interaction plots among parameters. For example, the top left subfigure shows the interaction between number of neuron in hidden layer 1 (Factor A) and number of neuron in hidden layer 2 (Factor B). As factor A and B has two levels each (low and high), the average over all momentum term and learning rate were calculated and plotted for the four possible combinations of A and B levels (i.e. (High A, High B), (High A, Low B), (Low A, High B) and (Low A, Low B)). This plot shows that the effect of the number of neurons in layer 2 on the average RMSE is small when the number of neurons in hidden layer 1 is at low level. As can be seen, Figure 2b) is very useful in interpreting significant interaction. However, it should not be used as the only mean to find the best setting of factors as the interpretation is subjective. As a result, in this study, the best setting of factors were determine from the setting that provide lowest average RMSE (Table 3) and the best setting in this case was 10 neurons in the first hidden layer, 6 neurons in the second layer, 0.1 learning rate and 0.8 momentum term.

The results from the optimized network can be used to calculate Curie temperature based on the given cumulant data. The extensive of the prediction are given in Ref [16], where accurate Curie temperature is obtained which strongly confirms the functionality the ANNs in providing a successful alternative approach in modeling ferromagnetic Curie temperature.

4. Conclusions and Suggestions

This paper has described techniques for the training of ANNs by using DoE. The result from the case study of the modeling of ferromagnetic materials property suggested that DoE approach could be successfully used to optimize back-propagation ANNs parameters. The factors that are found to be significant in the case study were number of hidden neurons in hidden layer 1 and 2, learning rate and momentum term. However, in the authors' opinion, the optimum setting of ANNS parameters are largely problem-dependent. Ideally and optimization process should be performed for each ANNs application, as the significant factors might be different for ANNs trained for different purpose.

5. Acknowledgements

The authors gratefully acknowledge the financial support from the Thailand Research Fund (TRF) and The Commission on Higher Education (Thailand).

References

[1] Udo G.J., Neural Networks applications in Manufacturing Processes, *Comput. Ind. Eng.*, 1992; 23: 97-100.

[2] Udo G.J., and Gupta Y.P., Applications of Neural Networks in Manufacturing Management Systems, *Prod. Planning Control,* 1994; 5: 258-270.

[3] Zhang H.C., and Huang, S.H., Applications of Neural Networks in Manufacturing: A State-of-the-art Survey, *Int. J. Prod. Res.*, 1995; 33: 705-728.

[4] Vellido A., Lisboa P.J.G., and Vaughan J., Neural Networks in Business: A Survey of applications (1992-1998), *Expert Syst. Appl.*, 1999; 17: 51-70.

[5] Wong B.K., Lai V.S., and Lam J., A Bibliography of Neural Network Business Applications research: 1994-1998, *Comput. Oper. Res.*, 2000; 27: 1045-1076.

[6] Gurney K., *An Introduction to Neural Networks*, London: UCL Press Limited, 1997.

[7] Box G.E.P., Hunter W.G., and Hunter, S., *Statistics for Experimenters: An Introduction to Design Data Analysis, and Model Building*, Wiley, New York, 1987.

[8] Montgomery D.C., *Design and Analysis of Experiments*, New York: John Wiley & Sons, 2001.

[9] Sukthomya W., and Tannock J.D.T., The Optimisation of Neural Network Parameters using Taguchi Design of Experiments Approach: An Application in Manufacturing Process Modeling, *Neural Comput. Appl.*, 2005; 14: 337-344.

[10] Kim Y.S., and Yum B.J., Robust Design of Multilayer Feed Forward Neural Networks: An Experimental Approach, *Appl. Artif. Intell.*, 2004; 17: 249-263.

[11] Packianather M.S., Drake P.R., and Rowland, H., Optimizing the Parameters of Multilayered Feed Forward Neural Networks through Taguchi Design of Experiments, *Qual. Reliab. Eng. Int.*, 2000; 16: 461-473.

[12] Yang S.M., and Lee G.S., Neural Network Design by using Taguchi Method, *J. Dyn. Syst. Meas. Contr.* 1999; 121: 560-563.

[13] Khaw J.F.C., Lim B.S., and Lim L.E.N., Optimal Design of Neural Networks using the Taguchi Method, *Neuralcomputing*, 1995; 7: 225-245.

[14] Binder K., Finite-size Scaling Analysis of Ising Model Block Distribution Functions, *Z. Phys. B*, 1981; 43: 119-140.

[15] McCoy B.M., and Wu T.T., *The Two–Dimensional Ising Model,* Cambridge: Harvard University Press, 1973.

[16] Laosiritaworn W.S., Application of Artificial Neural Network to Calculate Curie Temperature of Ferromagnetic Materials, *Adv. Mater. Res,* 2008; 55-57: 901-904.