ARTIFICIAL NEURAL NETWORK MODELLING USING BAYESIAN APPROACH IN GEOTECHNICAL ENGINEERING APPLICATIONS

Dr.D.Padmini

Associate Professor in Civil Engineering

Government College of Technology, Coimbatore, TN, India

ABSTRACT

Owing to the sensitive nature of the variables involved in geotechnical engineering problems, which is difficult to quantify and involve considerable uncertainty, the existing solutions e.g. mathematical, numerical and empirical, lack precision. The artificial neural networks have the ability to learn from data and model the non-linear relationship between the variables and hence, it has been widely used for modeling problems in geotechnical engineering. Several algorithms are available for developing artificial neural network models for regression problems among which, the backpropagation algorithm is widely used. This paper discusses the limitations of BPA algorithm and the efficient alternate algorithm that can be used for prediction for modeling geotechnical engineering problems.

Key words: ANN, Data division, Bayesian regularization.

1 INTRODUCTION

The data driven approaches are gaining popularity in geotechnical engineering as is evidenced by more number of applications over the last decade. These techniques have proved to be an efficient alternative to traditional methods for both estimation and forecasting. This follows from the fact that the theoretical methods have been simplified in order to model the complex and uncertain behaviour of the soil. But, the data driven approaches do not make any assumptions and their efficiency depends on the accuracy of the data.

Artificial neural network (ANN) is the most widely used method among the data driven approaches. It is a system composed of many simple processing elements operating in parallel. One of the most important properties of the ANNs is that they are universal approximators, i.e. they can fit any nonlinear function to any arbitrary degree of accuracy. This approach is usually accurate when numerical data are precise enough and representative of the system behavior.

The data division forms an important part of developing ANN model in which the available data are divided into training and testing sets. A brief description of ANN modeling process and the advantage of using Levenberg-Marquardt backpropagation algorithm (LMBPA) with Bayesian regularisation in data division for ANN model development are discussed in this paper.

2 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks referred to as parallel distributed, connectionist networks are ensembles of interconnected, usually nonlinear computational units called neurons or nodes, which are emulated by research into human brain system. ANN is a highly simplified model of a biological neural network when seen from a neurobiological

point of view. From mathematical point of view, they can be considered as a multivariate nonlinear modeling technique. The modeling philosophy is similar to a number of conventional statistical methods and they can be used to estimate functions from sample data as in the statistical technique. The major difference is that statistical approaches require guessing as to the functional dependency of outputs on inputs, whereas neural systems do not require articulation of any such physical or mathematical model (Agarwal et al 1994).

2.1 ARCHITECTURE OF ANN

ANNs can be designed as either feedforward or recurrent networks. The feedforward ANN has no feedback loops i.e. the neurons are connected only to those in the next layer. Also, they have the property that the outputs can be expressed as a deterministic function of the inputs and so the whole network represents a multivariate nonlinear functional mapping (Bishop 1995). The neurons in the recurrent network are fully or partially connected and they have feedback connections also. The feedforward ANN which is commonly referred to as multilayer perceptron (MLP) is most widely used in practice. The most important attribute of the MLP is that it can learn a mapping of any complexity. The network learning is based on repeated presentations of the training samples. The trained network often produce surprising results and generalisations in applications where explicit derivation of mapping or discovery of relationship is almost impossible (Zurada 2003). In MLP, the output function of the system is determined by network structure, connection strengths and the processing performed at computing elements or nodes. The network can be trained to learn and to perform a particular task. The ultimate goal for network training is to generalise, i.e. to have the output of the network equal to approximate target values, given inputs that are not in the training set. Generalisation performance is by far the most important criterion for ANN evaluation, particularly when accurate prediction is the main objective.

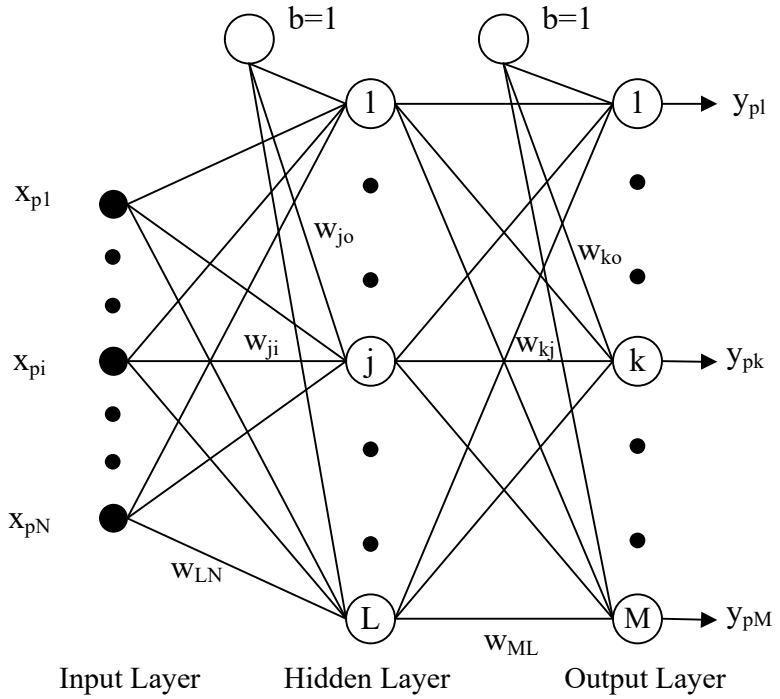


Figure 1 A typical Multilayer Feedforward Neural Network

A typical three layer MLP is shown in Figure 1. The Multilayer feedforward neural network (FNN) can have more than one hidden layer. However, a single hidden layer has been proven to be capable of providing accurate approximation to any complex nonlinear function provided there are sufficient hidden nodes (Cybenko 1989; Hornik et al 1989; Zhang et al 1998; Coulibaly et al 1999). It is the hidden nodes that allow the network to detect feature(s) in the data and to perform any nonlinear input-output mapping. There is no definite formula available to calculate the number of hidden nodes and usually it must be determined by trial and error method (Zhang 2007). The number of input and output nodes corresponds to the number of input and output variables in the problem. The number of nodes in the hidden layer and the type of algorithm used for training the network are problem-dependent and obtained by trial and error method.

2.2 ALGORTHM FOR MODEL DEVELOPMENT

Coulibaly et al (1999) reported that more than 23 learning rules have been proposed for training the network, however, none of them can guarantee the global minimum solution. Therefore, efficient network training is a challenging part of network design. A critical examination of literature indicates that more than 90% of the experiments make use of FNN trained by standard backpropagation algorithm (BPA), which is basically a gradient-based optimisation technique, developed by Rumelhart et al (1986). Strictly the notion of backpropagation refers to the reverse mode for computing the error gradient for a multilayer FNN using the chain rule (Werbos 1974). The detailed description of MLP trained using BPA can be found in many publications (Zurada 2003; Fausett 1994).

Although BPA training has proved to be efficient in a number of applications, it has inherent limitations of gradient-based techniques such as slow convergence and local search nature. Though various modifications have been proposed to the BPA, conventional second-order nonlinear optimisation methods such as the conjugate-gradient, the Levenberg-Marquardt and the quasi-Newton algorithms are usually faster than any variant of the BPA (Masters 1995; Hagan et al 1996) and achieve significant improvement in the forecast accuracy. The Levenberg-Marquardt backpropagation algorithm (LMBPA) is designed specifically for minimising a sum of squared error (Bishop 1995) and to overcome the limitations in the standard BPA.

The LMBPA uses the approximate Hessian matrix (second derivative of the error function E) in the weight update procedure as follows:

$$w_{new} = w_{old} - (Z^T Z + \lambda I)^{-1} Z^T \varepsilon(w_{old}) \qquad (1)$$

where I is the unit matrix, Z is the Jacobian matrix, $\mathbf{A} = Z^T Z$ is the Hessian matrix, λ is the small scalar variable which controls the learning process and ε is the residual error vector. When the scalar λ is zero, LM method is just Newton's method, using the approximate Hessian matrix and when λ is large, this becomes gradient-descent with a small step size. In this latter case, the step length is determined by λ^{-1}, so that it is clear that, for sufficiently large values of λ, the error will necessarily decrease since the equation (1) then generates a very small step in the direction of the negative gradient. In practice, a value must be chosen for λ and this value should vary appropriately during the minimisation process (Bishop 1995).

Building a parsimonious model with a minimum number of input variables and parameters to achieve high predictive accuracy without underfitting or overfitting problems is very much essential (Zhang 2007). Too many neurons in the hidden layer lead to overfitting i.e. the training data will be well modeled, but the network will be modeling the noise in the data as well as the trends. A network with an insufficient number of hidden nodes will have difficulty in learning data. That is, both too small and too large networks have poor prediction performance. Therefore, the network will not generalise well on the testing data. A common heuristic approach to avoid overfitting is early stopping in which the available data are divided into training, validation and testing sets. This approach involves monitoring the generalisation error for validation set and stopping training when the minimum validation error is observed. However, some care is needed when to stop training, since the validation error surface may have a local minima or long flat regions preceding a steep drop-off (Gori and Tesi 1992).

The LMBPA is usually used with early stopping method in which the performance of the networks is compared by evaluating the error function using an independent validation set and the network having the smallest error with respect to the validation set is selected. In practice, the available data is severely limited and it may not be possible to keep part of the data as validation set for model comparison purpose. To overcome these limitations, Mackay (1992) proposed the use of Bayesian backpropagation neural networks (BBPNN). The BBPNN is able to deal with the overfitting issue without using the validation data set.

The Bayesian approach allows different models (e.g. networks with different number of hidden nodes) to be compared using only the training data and it provides an objective and principled framework for dealing with the issues of model complexity which avoids many of the problems which arise when using the maximum likelihood.

The Bayesian approach enables the optimal weight decay parameters to be automatically during training (Mackay 1992; Bishop 1995). It provides a combined approach for dealing with issues of model complexity and overfitting. For regression problems, error bars or confidence intervals can be assigned to the prediction generated by the network.

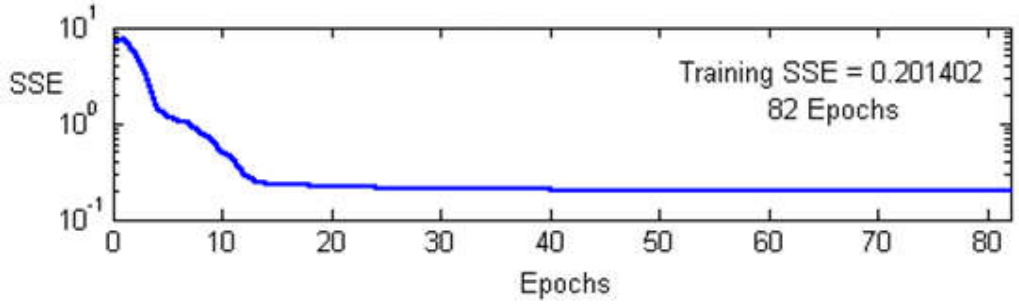
In the Bayesian frame work, the uncertainty in the weight space is assigned a probability distribution representing the degree of belief in the different values of the weight vector. This function is initially set to some prior distribution. Once the data has been observed, it can be converted to a posterior distribution through the use of Baye's theorem. By maximising the posterior distribution over the weights $\mathbf{w}$, the most probable parameter values $\mathbf{w}_{MP}$ can be obtained. Mackay (1992) has shown that maximising the posterior distribution corresponds to minimising the regularised error function. The posterior distribution is then used to evaluate the predictions of the trained network for new values of the input variables. The Bayesian approach allows the calculation of error bars on the network output, instead of just providing a single output.

The adjustment of the hyper parameters to their near-optimal values is carried out during training (Mackay 1992; Bishop 1995) and therefore eliminates the tedious and intensive task of searching for the optimal network for generalisation required by conventional backpropagation. The Bayesian framework essentially provides better generalisation and a statistical approach to deal with data uncertainty in comparison with the conventional backpropagation. The method has been successfully employed to analyse nonlinear multivariate problems (Bishop 1995).

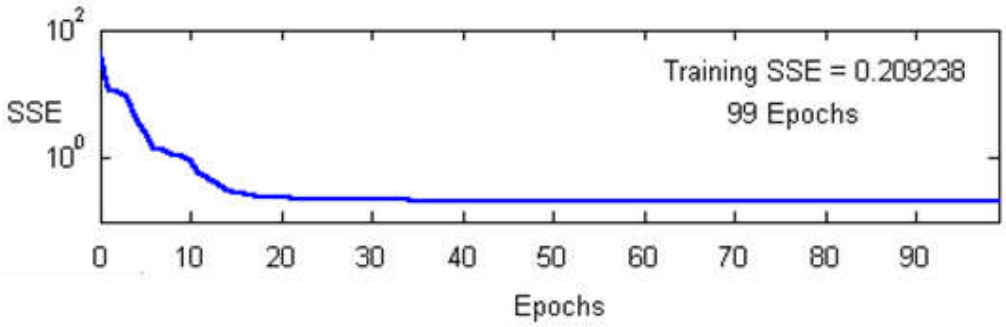
3 EXAMPLES

Padmini et al (2008) used the Levenberg-Marquardt backpropagation algorithm (LMBPA) with Bayesian regularisation for ANN model development to predict the ultimate bearing capacity of shallow foundations in cohesionless soils. The data used for calibrating and testing the ANN model were collected from literature, which include the load test data on small and large sized foundations, as well as the corresponding information regarding the footing and soil. The database has a total of 97 data sets, which consists the results of square, rectangular and strip footings of small and large sized foundations tested in sand beds of various densities. The data used are more evenly distributed (i.e. the number of data for large sized footings and smaller sized models are almost equal) that enhanced the performance of the ANN model.

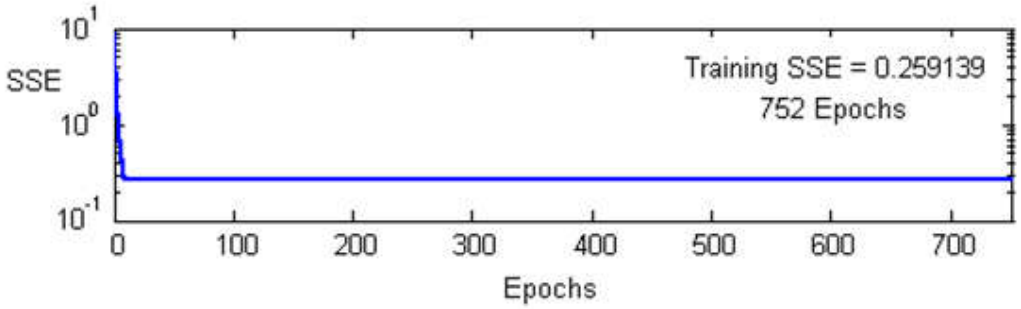
Tokar and Johnson (1999) have indicated that the way the data are divided can have a significant impact on the results obtained. Like all the empirical models, ANNs are unable to extrapolate beyond the range of their training data (Shahin et al 2002). Consequently, in order to develop the best possible model, with the available data, all the patterns that are contained in the data need to be included in the training set. The representative set of patterns for the training phase has been selected in such a way that it contains all the patterns including the maximum and the minimum values of all the input and output data. The true generalisation ability of the model can be tested, when all of the patterns in the training set are also part of the testing set. Therefore, it is essential that the data used for training and testing represent the same population (Masters 1993, as reported by Shahin et al 2002). In order to achieve this, several random combinations of training and testing sets in various proportions viz. 60% and 40 %, 65% and 35%, 70 % and 30 %, 75% and 25%, and 80% and 20% for training and testing respectively were tried in such a way that they represent the same population. On the basis of the above combinations, ANN models were developed. The variation of sum of squared error (SSE) with epochs for the five ANN models thus developed are shown in Figure 2.



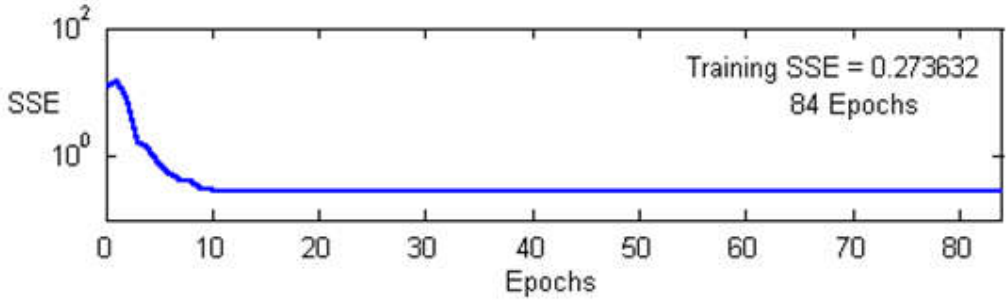
(a) 60% training and 40% testing



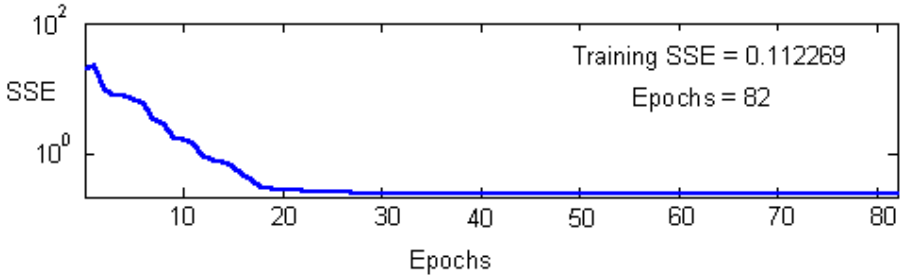
(b) 65% training and 35% testing



(c) 70% training and 30% testing



(d) 75% training and 25% testing



(e) 80% training and 20% testing

Figure 2 Epochs versus SSE for different partitioning of data

The MATLAB Toolbox for neural network was used for developing the above ANN models. Once training was successfully accomplished using the training set for each model, the performance of the models were evaluated on the test data. The performance of the developed ANN models for training and test data are indicated through performance indices namely Coefficient of correlation (r), Coefficient of efficiency (E), Root-mean-square error (RMSE), Mean absolute error (MAE), Mean bias error (MBE) in Table 1. The ANN model developed with Bayesian regularisation was able to predict well for all the above combinations. Out of the five combinations, the data division having 80 % data for training and 20 % data for testing resulted in the best performance.

Table 1 Performance statistics of ANN for the prediction of ultimate bearing capacity for different partitioning of data

(a) 60% (calibration) and 40% (testing) combination

Performance Index	Calibration (60%)	Testing (40%)
R	0.9874	0.9752
E	0.9740	0.9500
RMSE (kPa)	82.20	123.00
MAE (kPa)	57.20	81.40
MBE (kPa)	-4.76	-14.63
MARE (%)	20.71	21.71

(b) 65% (calibration) and 35 % (testing) combination

Performance Index	Calibration (65%)	Testing (35%)
R	0.9871	0.9765
E	0.9736	0.9582
RMSE (kPa)	80.40	124.90
MAE (kPa)	55.82	85.97
MBE (kPa)	-4.93	-10.38
MARE (%)	20.48	22.20

(c) 70% (calibration) and 30 % (testing) combination

Performance Index	Calibration (70%)	Testing (30%)
R	0.9842	0.9885
E	0.9680	0.9760
RMSE (kPa)	86.10	95.80
MAE (kPa)	57.93	68.50
MBE (kPa)	-5.07	-7.49
MARE (%)	19.54	20.03

(d) 75% (calibration) and 25 % (testing) combination

Performance Index	Calibration (75%)	Testing (25%)
R	0.9840	0.9880
E	0.9680	0.9760
RMSE (kPa)	85.30	99.90
MAE (kPa)	58.00	69.00
MBE (kPa)	-5.06	-6.11
MARE (%)	20.35	17.59

(e) 80% (calibration) and 20 % (testing) combination

Performance Index	Calibration (80%)	Testing (20%)
R	0.9950	0.9920
E	0.9890	0.9830
RMSE (kPa)	52.90	77.20
MAE (kPa)	39.98	57.02
MBE (kPa)	-1.78	-12.04
MARE (%)	16.64	15.43

Padmini et al (2008) have tested the potential of Artificial Neural Networks trained using Levenberg-Marquardt algorithm in predicting the pull out capacity of circular anchors in cohesionless soils. In order to avoid over fitting, Bayesian regularisation technique has been applied. The goodness-of-fit statistics for the developed ANN model such as coefficient of correlation (R), the coefficient of efficiency (E), root-mean-square error (RMSE) between the actual and predicted values, the mean bias error (MBE), and mean absolute relative error (MARE) are 0.998, 0.997, 3.758 kN, 0.132 kN and 7.46 % respectively which indicates the superior performance of ANN in the prediction of pull out capacity of circular anchors in sand using the above algorithm.

Goh and Chua (2013) demonstrated the efficiency of the evolutionary Bayesian back-propagation (EBBP) to locate the minima and handle the uncertainties in data for the following complicated nonlinear problems

1) Estimation of the Skin Friction capacity of driven piles in cohesive soils. 65 data records of driven pile load tests taken from the literature were used to assess the skin friction , out of which 45 and 20 randomly selected patterns were used for training and testing. The results indicated that the predictions using the EBBP were an improvement over those by the conventional method by Semple and Rigden method (1986).

2) To model the Pile Skin Friction for drilled shafts.

A total number of 127 field load test data patterns were used, of which 85 patterns were randomly selected as the training data and 42 patterns as the testing data. The regression equation given by Chen and Kulhawy (1994) was used for comparison. The predicted results were better than results obtained from the above regression equation particularly for training data.

3) Retaining wall deflection

The model was developed using 3844 training and 3081 testing data obtained from finite element method (FEM) analysis. 35 input parameters were used and wall deflection was the output. The EBBP predictions were in good agreement with the actual (FEM) all deflections.

4 CONCLUSIONS

The limited data set constrained us not to allocate any dataset for validation. However this limitation was taken care by using Bayasian regularization. Though the data set is limited, the application of the LMBPA with Bayesian regularisation resulted in good prediction accuracy, which encourages to apply this approach to other problems in geotechnical engineering as well where we may have limited data available.

ACKNOWLEDGEMENT

The author sincerely thanks Dr.K.Ilamparuthi, former Professor of Civil Engineering, College of Engineering, Anna University, Chennai for the valuable guidance while preparing this paper.

REFERENCES

1. Agarwal G., Weeraratne S. and Khilnani K. (1994), 'Estimating Clay Liner and Cover Permeability Using Computational Neural Networks', Proceedings of First Congress on Computing in Civil Engineering, Washington.

2. Bishop C.M. (1995), 'Neural Networks for Pattern Recognition', Oxford University Press, New Delhi.

3. Zurada J.M. (2003), 'Introduction to Artificial Neural Systems', Jaico Publishing House, New Delhi.

4. Cybenko G. (1989), 'Approximation by Superposition of a Sigmoidal Function', Mathematical Control Signals Systems, Vol.2, No.4, pp.303-314.

5. Hornik K., Stinchcombe M. and White H. (1989), 'Multilayer Feed Forward Networks are Universal Approximators', Neural Networks, Vol.2, No.5, pp.359-366.

6. Zhang G., Patuwo B.E. and Hu M.Y. (1998), 'Forecasting with Artificial Neural Networks: The State of the Art', International Journal of Forecasting, Vol.14, No.1, pp.35-62.

7. Coulibaly P., Anctil F. and Bobée B. (1999), 'Prévision Hydrologique par Réseaux de Neurones Artificiels: état de l'art', Canadian Journal of Civil Engineering, Vol.26, No.3, pp.292-304.

8. Zhang G.P. (2007), 'Avoiding Pitfalls in Neural Network Research', IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, Vol.37, No.1, pp.3-16.

9. Rumelhart D.E., Hinton G.E. and Williams R.J. (1986), 'Learning Internal Representations by Error Propagation', Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Edited by Rumelhart D.E. and McClelland J.L., MIT press: Cambridge, MA, Vol.1, pp.318-362.

10. Werbos P. (1974), 'Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences', Ph.D. Dissertation, Committee on Applied Mathematics, Harvard University, Cambridge, MA.

11. Fausett L.V. (1994), 'Fundamentals of Neural Networks:Architectures, Algorithms and Applications', Prentice-Hall, Englewood Cliffs, NJ.

12. Masters T. (1995), 'Advanced Algorithms for Neural Networks: A C++ Sourcebook', John Wiley and Sons, New York.

13. Hagan M.T., Demuth H.B. and Beale M. (1996), 'Neural Network Design', PWS Publishing Company, Boston, MA.

14. Gori M. and Tesi A. (1992), 'On the Problem of Local Minimum in Backpropagation', Transactions on Pattern Analysis and Machine Learning, IEEE, Vol.14, No.1 pp.76-86.

15. MacKay D.J.C. (1992), 'Bayesian Interpolation', Neural Computation, Vol.4, No.3, pp.415-447.

16. Padmini D., Ilamparuthi K. and Sudheer K.P. (2008), 'Ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neurofuzzy models', Computers and Geotechnics, Elsevier, Vol. 35, No.1, pp. 33-46 (peer reviewed).

17. Tokar S.A. and Johnson P.A. (1999), 'Rainfall-Runoff Modeling using Artificial Neural Networks', Journal of Hydrologic Engineering, Vol.4, No.3, pp.232-239.

18. Shahin M.A., Maier H.R. and Jaksa M.B. (2002), 'Predicting Settlement of Shallow Foundations using Neural Networks', Journal of Geotechnical and Geoenvironmental Engineering, ASCE, Vol.128, No.9, pp.785-793.

19 Padmini D., Ilamparuthi K. and Sudheer K.P. (2008), 'Pullout capacity prediction of circular plate anchors in cohesionless soils using artificial neural networks', 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG), Goa, pp. 1818-1826.

20 Goh A.T.C. and Chua C.G. (2013), 'Geotechnical Applications of Bayesian Neural Networks', In book: Metaheuristics in Water, Geotechnical and Transport Engineering, pp. 271-287.