Computer Assisted Mechanics and Engineering Sciences, **12**: 259–265, 2005.

Prediction of concrete fatigue durability using Bayesian neural networks

Marek Słoński
Cracow University of Technology, Institute of Computer Methods in Civil Engineering
ul. Warszawska 24, 31-155 Kraków, Poland

(Received March 31, 2005)

The utility of Bayesian neural networks to predict concrete fatigue durability as a function of concrete mechanical parameters of a specimen and characteristics of the loading cycle is investigated. Bayesian approach to learning neural networks allows automatic control of the complexity of the non-linear model, calculation of error bars and automatic determination of the relevance of various input variables. Comparative results on experimental data set show that Bayesian neural network works well.

Keywords: Bayesian neural networks, concrete fatigue durability, prediction

1. Introduction

One of the most straightforward approaches to a problem of data modelling is to represent the underlying model structure in parametric form. The optimal values of the parameters can then be found to give the best fit to the data. Neural networks can be considered to be parametric modelling method, that allow to build very flexible models. In some data analysis problems more complex models usually fit the data better but may fit the noise rather than the underlying functional relationship (problem known as a overfitting). Such models therefore give poor performance on unseen data.

In the Bayesian approach complexity control of the model can be handled consistently. The unknown complexity of the neural network model is represented by priors for the hyperparameters and the resulting prediction is averaged over all models weighted by their posterior probability given the data. For neural networks Bayesian methods were introduced among others by MacKay and Neal [8, 11].

Bayesian neural networks have proved to be an effective tool for regression problems. In paper [7] Bayesian neural network was applied to prediction of concrete properties and the Bayesian approach gave better results than alternative non-Bayesian methods in the case problem. Bayesian neural network was also used to prediction of deformed and annealed microstructures [1] and in fault identification in cylinders using vibration data [9].

In previous works various feed-forward neural network models were used to the problem of predicting concrete fatigue durability: back propagation (BPNN) [5], radial basis function (RBFNN) [6, 15], adaptive neuro-fuzzy inference system (ANFIS) [16], and fuzzy weights NNs (FWNN) [14].

In this paper a Bayesian neural network was applied to the problem of predicting the concrete fatigue durability. Concrete durability is defined as the limit number of cycles N which causes the specimen fatigue damage. As a reference method an early-stopped committee of 10 MLP networks was used. Early-stopped committee is a simple and fast method and has proved to be quite a robust method [7].

2. Bayesian neural networks

In this section the application of Bayesian inference to neural network learning is considered. Standard neural networks training methods give a single set of values of network parameters $\mathbf{w}^*$ as a result of error function $E(\mathbf{w})$ minimization. By contrast, Bayesian approach considers a probability distribution over parameters space. A *prior* probability distribution $p(\mathbf{w})$ is first defined, which expresses beliefs about the parameters before the data is observed. Once the data are observed, Bayes's theorem is used to update beliefs and the *posterior* probability distribution $p(\mathbf{w}|\mathcal{D})$ is obtained

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}, \tag{1}$$

where $p(\mathcal{D}|\mathbf{w})$ is the likelihood of parameters $\mathbf{w}$, which represents a model for the noise process on the target data and $p(\mathcal{D})$ is a normalizing constant which is written as

$$p(\mathcal{D}) = \int p(\mathcal{D}|\mathbf{w})p(\mathbf{w})\mathrm{d}\mathbf{w}. \tag{2}$$

The *posterior* probability distribution is then used to evaluate the predictions of the trained network for new input vector. The predicted distribution $p(t_{n+1}|\mathbf{x}_{n+1}, \mathcal{D})$ of target values t_{n+1} for a new input vector $\mathbf{x}_{n+1}$ once the data set $\mathcal{D} = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_n, t_n)\}$ was observed, is expressed as an integration over the *posterior* distribution of weights in the form

$$p(t_{n+1}|\mathbf{x}_{n+1}, \mathcal{D}) = \int p(t_{n+1}|\mathbf{x}_{n+1}, \mathcal{D}, \mathbf{w})p(\mathbf{w}|\mathcal{D})\mathrm{d}\mathbf{w}, \tag{3}$$

where $p(t|\mathbf{x}, \mathcal{D}, \mathbf{w})$ is the conditional probability model [8, 4, 2].

For regression problems, it is generally assumed that the measured values (target variable) t_n will contain noise e_n, so the model's prediction, $f(\mathbf{x}_n; \mathbf{w})$, is related to the target output by

$$t_n = f(\mathbf{x}_n; \mathbf{w}) + e_n. \tag{4}$$

The commonly used noise model for the regression problems is a Gaussian noise model $N(0, \sigma^2)$ with zero mean and constant variance σ^2. From (4) the probability of observing a data value t for a given input vector $\mathbf{x}$ is then given by

$$p(t|\mathbf{x}, \mathbf{w}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(f(\mathbf{x}; \mathbf{w}) - t)^2}{2\sigma^2}\right\}, \tag{5}$$

where $f(\mathbf{x}; \mathbf{w})$ represents a network function as the mean of distribution and parameter σ^2 controls the variance of the noise. For data points independently drawn from distribution defined by Eq. (5) the likelihood of the network parameters given data set $\mathcal{D}$ is

$$p(\mathcal{D}|\mathbf{w}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\sum_{n=1}^{N} \frac{(f(\mathbf{x}_n; \mathbf{w}) - t_n)^2}{2\sigma^2}\right\}. \tag{6}$$

In the paper a Bayesian neural network with one hidden layer was applied to model the relationship between the inputs and output variables. In matrix format this model can be written as

$$f(\mathbf{x}; \mathbf{w}) = b_o + \mathbf{w}_o S(\mathbf{b}_h + \mathbf{w}_h \mathbf{x}), \tag{7}$$

where $\mathbf{w}$ denotes all the parameters $\mathbf{w}_h$, $\mathbf{b}_h$, $\mathbf{w}_o$ and b_o, which are the hidden layer weights and biases, and the output layer weights and bias, respectively. A function $S(z) = \tanh(z)$ is the hidden node activation function.

For the model parameters $p(\mathbf{w})$ the Gaussian distribution $N(\mathbf{0}, \alpha_k)$ was assumed, where the α's are the inverse variance hyperparameters ($k = w_h, b_h, w_o, b_o$). For example, Gaussian prior distribution for the hidden layer weights $p(\mathbf{w}_h)$ is defined as follows

$$p(\mathbf{w}_h) = \frac{1}{Z_W(\alpha_{w_h})} \exp\left(-\frac{\alpha_{w_h}}{2}\|\mathbf{w}_h\|^2\right), \tag{8}$$

where $Z_W(\alpha_{w_h}) = \left(\frac{2\pi}{\alpha_{w_h}}\right)^{W_h/2}$ and W_h is the number of the hidden layer weights.

Also a hierarchical Gaussian prior distribution for the hidden layer weights called Automatic Relevance Determination (ARD) was applied [8, 10]. The ARD prior distribution is an automatic method for determining the relevance of the inputs. In ARD all weights connected to the same input i has the same variance hyperparameters α_i. These hyperparameters are important because they control the complexity of the model. The irrelevant inputs should have smaller weights in the connections to the hidden units than more important weights [13, 7].

In general, the required integrations in Eq. (3) are analytically intractable. There are some techniques to solve this problem. The integration can be performed by making simplifying assumptions about the form of $p(\mathbf{w}|\mathcal{D})$ (a local Gaussian approximation known as the Laplace approximation), by using variational methods or by employing Monte Carlo methods, of which the most successful is that of hybrid Monte Carlo [11, 10, 17]. In the Monte Carlo approach the integrals are approximated by finite sums of the form:

$$\int p(t|\mathbf{x}, \mathcal{D}, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) d\mathbf{w} \approx \frac{1}{m} \sum_{i=1}^{m} p(t|\mathbf{x}, \mathcal{D}, \mathbf{w}_i), \tag{9}$$

where $\mathbf{w}_i$ are samples of weight vectors generated from distribution $p(\mathbf{w}|\mathcal{D})$.

3. Concrete fatigue durability prediction

In this section a problem of concrete fatigue durability prediction is defined and some related works on solving this problem are presented. The concrete fatigue durability is defined as a function of the concrete compressive strength (f_c), ratio of minimal and maximal strength in compressive cycle of loading ($R = \sigma_{\min}/\sigma_{\max}$), frequency of the loading cycle (f) and ratio of compressive fatigue strength of concrete ($\chi = f_{cN}/f_c$). The problem of predicting concrete fatigue durability was formulated as a mapping from the input vector $\mathbf{x}_{(4x1)} = \{fc, R, f, \chi\}$ to the scalar output $y = \log N$.

In order to train and test Bayesian neural network model a large amount of representative experimental data is required. In paper [3] a wide experimental evidence was described and compiled, corresponding to more than 400 tests performed in 14 laboratories. The concrete specimens were subjected to cycles of compressive loadings and the numbers of cycles N which caused the specimens fatigue damage were measured. In this paper 216 selected results of laboratory tests from 8 papers collected in [3] was used. Table 1 shows the basic statistical parameters of analysed data used in modelling the relation.

Table 1. Input and output variables statistical details

Variable	Min	Max	Mean	St. Dev.
f_c [MPa]	20.70	45.20	34.68	8.84
R [-]	0.00	0.88	0.14	0.18
f [Hz]	0.025	150.0	21.30	39.38
χ [-]	0.49	0.94	0.74	0.11
$\log N$ [-]	1.86	7.34	4.56	1.41

In paper [3] an empirical formula was derived by Furtak as the following implicit relation between variables:

$$\log N = \frac{1}{A}\left[\log(1.16 \cdot C_f/\chi) + \log(1 + B \cdot R \cdot \log N)\right], \tag{10}$$

where: $\chi = f_{cN}/f_c$ and $R = \sigma_{\min}/\sigma_{\max}$ and the parameters according to paper [3] have the following values: $A = 0.008 - 0.118 \cdot \log(\sigma_I/f_c)$, $B = 0.118 \cdot (\sigma_{II}/\sigma_I - 1)$, $C_f = 1 + 0.07 \cdot (1 - R) \cdot \log f$, σ_I and σ_{II} are critical strengths.

4. Experiments

A Bayesian neural network with a single hidden layer of ten hyperbolic tangent units (neurons) and linear output unit was used to model the relationship between the inputs and the output. The network weights were initially randomised from a Gaussian distribution. The Bayesian neural network with ARD was trained using Hybrid Monte Carlo (HMC) implemented in Flexible Bayesian Modelling (FBM) software by Radford Neal [11, 12].

The early-stopped committee of MLP networks (ESC MLP) model contained 10 networks which were created with different division of the training data into learning (estimating) and stopped (validation) sets for each member. One third of the training examples (rounded down if necessary) were used for stopping (validation) and the rest for learning (estimating the weights). 100 iterations of the scaled conjugate gradient algorithm was used to optimise the network weights. The simplest form of committee involves taking the output of the committee to be the average of the outputs of L networks. The committee prediction [2] is defined as $f_{\text{com}}(\mathbf{x}) = \frac{1}{L}\sum_{i=1}^{L} f_i(\mathbf{x})$.

ESC MLP network model was implemented and trained by scaled conjugate gradient optimization method in MATLAB with Netlab toolbox [10].

The experiments were performed using 216 examples (patterns). The generalization capability of the models in predicting concrete fatigue durability was estimated by cross-validation metod. To reduce the effects of data partitioning on the generalisation performance, the models were evaluated for ten different (random) partitions of the data (195 or 194 for training and 21 or 20 for testing).

In the analysis both the inputs and output variables were first standardized to zero mean and unit standard deviation by transformation

$$\tilde{x}_i^n = \frac{x_i^n - \bar{x}_i}{s_i}, \tag{11}$$

where $\bar{x}_i$ is an average value and s_i is the standard deviation

$$\bar{x}_i = \frac{1}{N}\sum_{n=1}^{N} x_i^n, \qquad s_i = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_i^n - \bar{x}_i)^2}. \tag{12}$$

With data normalization network weights can all be expected to have similar values if the inputs are equally important, and can be initialised randomly.

5. Results

The following results presented below are computed and plotted for the first fold split of data into training and testing sets. In Fig. 1 concrete fatigue durability values measured vs predicted by Bayesian neural network with 1σ error bars are shown.

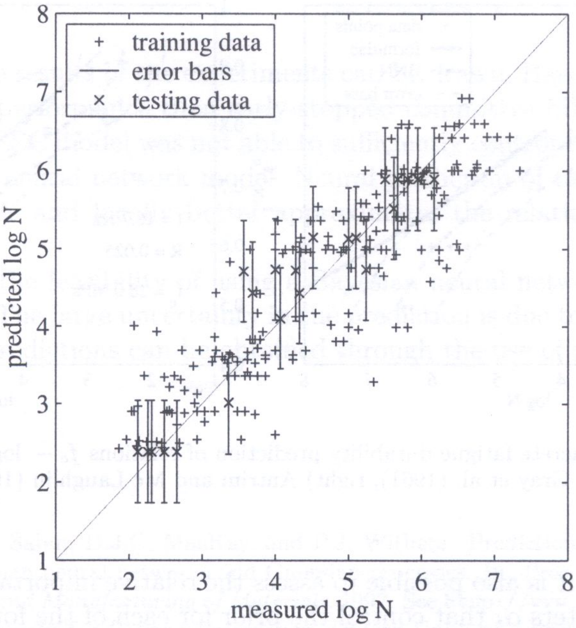

Fig. 1. Concrete fatigue durability values measured vs predicted by Bayesian neural network with 1σ error bars for 22 testing patterns

In Table 2 the following estimated prediction errors are presented:

- root mean square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(t_n - f(\mathbf{x}_n))^2}, \tag{13}$$

- average percentage error (APE)

$$\mathrm{APE} = \frac{1}{N}\sum_{n=1}^{N}|1 - f(\mathbf{x}_n)/t_n| \cdot 100\%, \tag{14}$$

where t_n is target value, $f(\mathbf{x}_n)$ is output value of a model f for the input vector $\mathbf{x}_n$. Both errors are values averaged over 10-fold cross-validation with standard deviation of the mean.

Table 2. Comparison of generalization performance of various models in predicting fatigue durability

Model	RMSE±std	APE±std [%]
Furtak's formulae	0.885 ± 0.11	18.6 ± 3.0
ESC NN model	0.731 ± 0.10	14.1 ± 2.6
BNN model	0.693 ± 0.09	13.3 ± 2.4

In Fig. 2 the relations $f_c - \log N$ for empirical formula by Furtak and Bayesian neural network are shown for tests performed by Gray et al. (1961) and for tests made by Antrim and Mc Laughlin (1959). The dash-dot lines in Fig. 2 show the relations modelled by Bayesian neural network and the dashed lines are the corresponding 1σ error bars which were computed on the basis of RMS error for all 216 patterns.

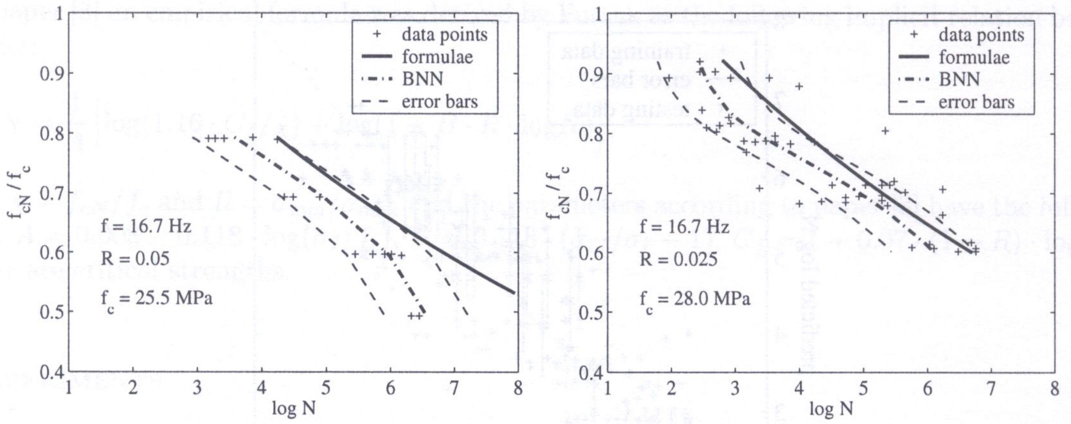

Fig. 2. Comparison of concrete fatigue durability prediction of relations $f_c - \log N$ for tests performed by: left) Gray et al. (1961), right) Antrim and Mc Laughlin (1959)

In Bayesian framework it is also possible to assess the relative importance of inputs by observing values of the hyperparameters α_i that control the prior for each of the four input variables. Table 3 gives mean and standard deviation for the hyperparameters associated with each input variable across each of the ten data partitions. A small value for α_i corresponds to a large variance prior which allows weights of large magnitude. Such a variable is very relevant for predicting the output. On the other hand the variable R has a larger value than the other three hyperparameters and can be identified as the least relevant variable in predicting concrete fatigue durability.

Table 3. Mean and standard deviation for the ARD hyperparameters values

	f_c	R	f	χ
$\bar{\alpha}$	12.85	22.14	1.44	2.45
s_α	24.78	36.54	1.73	2.04

Figure 3 shows the variation in mean square error on the standardized testing set for the networks sampled in the last 100 iterations. There is also shown the mean squared error on the testing set using the averaged predictions from all networks sampled up to the given iteration (within the last 100). There is clearly a benefit from averaging the predictions of some number of networks rather than making predictions from the single network trained with classical methods.

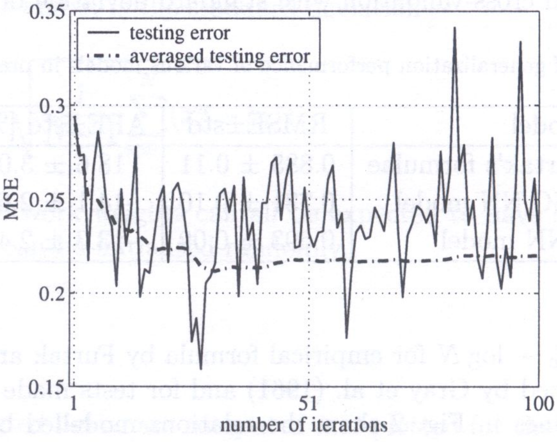

Fig. 3. Mean square error on the testing set for networks sampled in the last 100 iterations (one of the runs)

6. Conclusions

Some conclusions from the results of the experiments can be drawn. Bayesian neural network model has better generalisation performance than early-stopped committee MLP ESC model. The early-stopped committee MLP ESC model was not able to sufficiently control the complexity of the model comparing with Bayesian neural network model. Neural prediction of the number of fatigue cycles gives lower values of $\log N$ and locally better approximates the relation than estimation by the empirical formula.

The results confirmed the feasibility of using a Bayesian neural network to model the data and discover the relationship. The large uncertainty in the prediction is due to noise in the training data. Smaller uncertainties in predictions can be obtained through the use of a larger and more accurate data set.

References

[1] C.A.L. Bailer-Jones, T.J. Sabin, D.J.C. MacKay, and P.J. Withers. Prediction of deformed and annealed microstructures using Bayesian neural networks and Gaussian processes. In: *Proc. of the Australia-Pacific Forum on Intelligent Processing and Manufacturing of Materials.* 1997. See `http://www.mpia-hd.mpg.de/homes/calj/`.

[2] C.M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, Oxford, 1995. See `http://research.microsoft.com/~cmbishop/`.

[3] K. Furtak. Strength of the concrete under multiple repeated loads (in Polish). *Arch. of Civil Eng.*, **30**, 1984.

[4] D. Husmeier. *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Point Predictions.* Perspectives in Neural Computing. Springer London, 1999.

[5] J. Kaliszuk, A. Urbańska, Z. Waszczyszyn, and K. Furtak. Neural analysis of concrete fatigue durability on the basis of experimental evidence. *Arch. of Civil Eng.*, **38**, 2001.

[6] J. Kaliszuk and Z. Waszczyszyn. Application of a radial basis function neural network for prediction of concrete fatigue durability (in Polish). In: *Proc. 46-th Polish Civil Eng. Conf.*, volume 2, pages 193–198. Krynica-Wrocław, 2000.

[7] J. Lampinen and A. Vehtari. Bayesian approach for neural networks – review and case studies. *Neural Networks*, 14(3):7–24, April 2001. See `http://www.lce.hut.fi/~ave/`.

[8] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003. See `http://www.inference.phy.cam.ac.uk/mackay/itila/`.

[9] T. Marwala. Scaled conjugate gradient and bayesian training of neural networks for fault identification in cylinders. *Computers and Structures*, **79**(32): 2793–2803, December 2001. See `http://dept.ee.wits.ac.za/~marwala/`.

[10] I.T. Nabney. *Netlab: Algorithms for Pattern Recognition.* Springer-Verlag, London, 2002. See `http://www.ncrg.aston.ac.uk/netlab/`.

[11] R.M. Neal. Bayesian training of backpropagation networks by the hybrid Monte Carlo method. Technical Report CRG-TR-92-1, 1992. See `http://www.cs.toronto.edu/~radford`.

[12] R.M. Neal. *Bayesian Learning for Neural Networks.* Number 118 in Lecture Notes in Statistics. Springer New York, 1996.

[13] R.M. Neal. Assessing relevance determination methods using DELVE. In: C. M. Bishop, ed. 1998. See `http://www.cs.toronto.edu/~radford`.

[14] E. Pabisek, M. Jakubek, and Z. Waszczyszyn. A fuzzy neural network for the analysis of experimental structural mechanics problems. In: J. Kacprzyk and L. Rutkowski, eds. *Neural Networks and Soft Computing*, Advances in Soft Computing. Physica-Verlag (Springer), 2002.

[15] R. Putanowicz and Z. Waszczyszyn. Neural network identification of concrete properties. In: W. Duch, ed. *Proceedings of EANN'99.* Warszawa, 1999.

[16] M. Słoński. Application of ANFIS neuro-fuzzy system for prediction of concrete fatigue durability. In: *13-th Inter-Institute Seminar for Young Researchers.* Vienna, Austria, 2001. (Abstract).

[17] A. Vehtari, S. Särkkä, and J. Lampinen. On MCMC sampling in Bayesian MLP neural networks. In: Shun-Ichi Amari, C. Lee Giles, M. Gori, and V. Piuri, eds. *IJCNN'2000: Proc. of the 2000 International Joint Conference on Neural Networks*, volume I, pages 317–322. IEEE, 2000. See `http://www.lce.hut.fi/~ave/`.