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
e-mail: mslonski@twins.pk.edu.pl

Abstract

The utility of Bayesian MLP 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

Bayesian methods are a robust and powerful approach to prediction problems and can be incorporated into the neural network approach. For neural networks Bayesian methods were introduced among others by MacKay and Neal [7, 9]. Standard neural networks training methods give a single "optimal" vector of network parameters $\boldsymbol{w}^*$ as a result of cost function $E(\boldsymbol{w})$ minimization [2]. In the Bayesian approach network parameters are treated as random variables. Thus Bayesian methods for neural networks yield posterior distribution of network parameters.

Bayesian neural networks have proved to be an effective tool for regression problems. In paper [6] 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 and Gaussian process models were also used to prediction of deformed and annealed microstructures [1].

In this paper we have applied a Bayesian neural network to the problem of predicting the concrete fatigue durability which is defined as the limit number of cycles N which causes the specimen fatigue damage. As a reference method we have used an early-stopped committee of 10 MLP networks. Early-stopped committee is an *ad hoc* method but it is fast and has proved to be quite a robust method when used as committee of early-stopped MLPs [6].

2. Bayesian neural networks

In the Bayesian approach, we first define a *prior* probability distribution $p(\boldsymbol{w})$ which expresses our beliefs about the parameters before the data is observed. Once we observe the data, Bayes's theorem can be used to update our beliefs and we obtain the *posterior* probability density $p(\boldsymbol{w}|\mathcal{D})$.

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

$$f(\boldsymbol{x};\boldsymbol{w}) = b_o + \boldsymbol{w}_o g(\boldsymbol{b}_h + \boldsymbol{w}_h \boldsymbol{x}), \tag{1}$$

where $\boldsymbol{w}$ denotes all the parameters $\boldsymbol{w}_h$, $\boldsymbol{b}_h$, $\boldsymbol{w}_o$ and b_o, which are the hidden layer weights and biases, and the output layer weights and bias, respectively. The function $g(\cdot)$ is tanh activation function.

2.1. Gaussian noise model

In general, the measured values t will contain noise e, so the model's prediction, $f(\boldsymbol{x};\boldsymbol{w})$, is related to the target output by

$$t = f(\boldsymbol{x};\boldsymbol{w}) + e. \tag{2}$$

The commonly used noise model for the regression problems is a Gaussian $N(0,\sigma^2)$ with zero mean and constant variance σ^2.

The probability of observing a data value t for a given input vector $\boldsymbol{x}$ is then given by

$$p(t|\boldsymbol{x},\boldsymbol{w}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(f(\boldsymbol{x};\boldsymbol{w})-t)^2}{2\sigma^2}\right\} \tag{3}$$

where $f(\boldsymbol{x};\boldsymbol{w})$ represents a network function as the mean of distribution and parameter σ controls the variance of the noise.

We also have used as a noise model Student's t-distribution with unknown degrees of freedom ν.

2.2. Prior distribution of model parameters

We have assumed prior distribution for the model parameters $p(\boldsymbol{w})$ to be Gaussian distribution $N(0,\alpha_k)$, 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 $p(\boldsymbol{w}_h)$ is defined as follows

$$p(\boldsymbol{w}_h) = \frac{1}{Z_W(\alpha_{w_h})} \exp\left(-\frac{\alpha_{w_h}}{2}\|\boldsymbol{w}_h\|^2\right), \tag{4}$$

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.

We have also applied a hierarchical Gaussian prior distribution for the hidden layer weights called Automatic Relevance Determination (ARD) [7, 8]. 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 [6].

2.3. Posterior distribution of model parameters

Bayesian neural network posterior distribution of network parameters given the data $\mathcal{D}$ from Bayes' theorem:

$$p(\boldsymbol{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\mathcal{D})}, \tag{5}$$

where $p(\mathcal{D}|\boldsymbol{w})$ is the likelihood of parameters $\boldsymbol{w}$, $p(\boldsymbol{w})$ is the prior distribution and $p(\mathcal{D})$ is a normalizing constant.

For data points independently drawn from distribution defined by equation (3) the likelihood of the network parameters given data set $\mathcal{D}$ is

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

2.4. *Distribution of network output*

The predicted distribution $p(t_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D})$ of target values t_{n+1} for a new input vector $\boldsymbol{x}_{n+1}$ once we have observed the data set $\mathcal{D} = \{(\boldsymbol{x}_1, t_1), \ldots, (\boldsymbol{x}_n, t_n)\}$ can be expressed as an integration over the posterior distribution of weights of the form

$$p(t_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}) = \int p(t_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, \boldsymbol{w}) p(\boldsymbol{w}|\mathcal{D}) d\boldsymbol{w}, \quad (7)$$

where $p(t|\boldsymbol{x}, \mathcal{D}, \boldsymbol{w})$ is the conditional probability model.

In general, the required integrations in equation (7) are analytically intractable. The integration can be performed by making simplifying assumptions about the form of $p(\boldsymbol{w}|\mathcal{D})$ or by employing Markov Chain Monte Carlo (MCMC) methods to evaluate the integrals numerically:

$$\int p(t|\boldsymbol{x}, \mathcal{D}, \boldsymbol{w}) p(\boldsymbol{w}|\mathcal{D}) d\boldsymbol{w} \approx \frac{1}{m} \sum_{i=1}^{m} p(t|\boldsymbol{x}, \mathcal{D}, \boldsymbol{w}_i) \quad (8)$$

where $\boldsymbol{w}_i$ are samples of weight vectors generated from distribution $p(\boldsymbol{w}|\mathcal{D})$. In the MCMC, samples are generated using a Markov chain. Its stationary distribution is the desired posterior distribution of weights.

3. Predicting concrete fatigue durability

In our analysis, the concrete fatigue durability is a function of concrete compressive strength. (f_c), ratio of minimal and maximal strength in compressive cycle of loading ($R = \sigma_{min}/\sigma_{max}$), ratio of compressive fatigue strength of concrete ($\chi = f_{cN}/f_c$) and frequency of the loading cycle (f). The problem of predicting concrete fatigue durability was formulated as a mapping from the input vector $\boldsymbol{x}_{(4x1)} = \{fc, R, f, \chi\}$ to the scalar output $y = \log N$.

3.1. *Data set*

In order to train and test Bayesian MLP 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 we have used 216 selected results of laboratory tests from 8 papers collected in [3].

Table 1 shows the inputs and output variables details (range, mean and standard deviation of data set used in modelling the relation). In the analysis both the inputs and output variables were first standardized to zero mean and unit standard deviation as follows:

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x} \quad (9)$$

where $\bar{x}$ is an average value and σ_x is the standard deviation:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (10)$$

Table 1. Input and output variables details

Variable	Range	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

$$\sigma_x = \sqrt{\frac{1}{1-n} \sum_{i=1}^{n} (x_1 - \bar{x})^2}. \quad (11)$$

3.2. *Related works*

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

$$\log N = \frac{1}{A} \Big[\log(1.16 \cdot C_f/\chi) + \log(1 + B \cdot R \cdot \log N) \Big] \quad (12)$$

where: $\chi = f_{cN}/f_c$ i $R = \sigma_{min}/\sigma_{max}$ 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.

In previous works various feed-forward neural network models were used to the problem of predicting concrete fatigue durability: back propagation (BPNN) [4], radial basis function (RBFNN) [5, 11], adaptive neuro-fuzzy inference system (ANFIS) [12], and fuzzy weights NN (FWNN) [10].

4. Experiments and results

4.1. *Models settings used in the experiments*

We have used Bayesian MLP neural networks models with a single layer of hyperbolic tangent units (neurons) and linear output units. All units had biases. The number of hidden units was either 5 or 10. Bayesian neural network models were trained with MCMC (Markov Chain Monte Carlo) method.

Bayesian neural network (BNN) was compared with two models: an early-stopped committee of MLP networks (ESC MLP) and Gaussian Process (GP) model, which is a non-parametric regression method [1]. The 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. We have used one third of the training examples (rounded down if necessary) for stopping (validation) and the rest for learning (estimating the 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 is in the form [2]

$$f_{com}(\boldsymbol{x}) = \frac{1}{L} \sum_{i=1}^{L} f_i(\boldsymbol{x}). \quad (13)$$

Bayesian neural network and Gaussian Process models were implemented and trained with MCMC (Markov Chain Monte Carlo) method in Flexible Bayesian Modelling (FBM) software by Radford Neal [9] (original programs are available from Radford Neal's web page http://www.cs.utoronto.ca/∼radford/).

Table 2. Comparison of generalization performance of various models in predicting concrete fatigue durability. The shown testing errors give the RMS errors and average percentage errors (APE) averaged over 10-fold cross-validation with standard deviation of the mean.

Model	RMSE±std	APE±std [%]
Furtak's form.	0.885 ± 0.11	18.6 ± 3.0
ESC 4-5-1	0.703 ± 0.10	13.5 ± 2.6
BNN ARD N	0.694 ± 0.09	13.3 ± 2.3
BNN – t_ν	0.692 ± 0.09	13.0 ± 2.5
MLP 4-10-1	0.727 ± 0.17	14.2 ± 3.8
ESC 4-10-1	0.731 ± 0.10	14.1 ± 2.6
BNN ARD N	0.693 ± 0.09	13.3 ± 2.4
BNN – t_v	0.689 ± 0.09	13.0 ± 2.6
GP – t_v	0.674 ± 0.09	13.0 ± 2.6

ESC MLP network model was implemented and trained by scaled conjugate gradient optimization method in MATLAB with Netlab toolbox [8] (available from web page http://www.ncrg.aston.ac.uk/netlab/).

4.2. *Results of the experiments*

The generalization capability of the models in predicting concrete fatigue durability was estimated by 10-fold cross-validation. The estimated prediction errors (mean and standard deviations of the RMSE and APE (average percentage error) are presented in Table 2.

In column *Model* ARD means that the hierarchical Gaussian prior distribution for the weights was used and '–' means that ARD was not used. The letter *N* indicates that Gaussian noise model was assumed and t_ν means that the Student's t-distribution with unknown degrees of freedom ν noise model was used.

In Fig. 1 the relations $f_c - \log N$ for empirical formula by Furtak [3] and Bayesian neural network are shown for tests performed by Gray et al. (1961) and for tests made by Antrim and Mc Laughlin (1959) taken from [3] (for the first fold split of data). The dash-dot lines in Fig. 1 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 (for the first fold split of data).

In Fig. 2 concrete fatigue durability values measured vs predicted by Bayesian neural network (for the first fold split of data) with 1σ error bars (for 22 testing patterns only) are shown.

Fig. 3 shows the variation in mean square error on the standardized training and testing sets (for the first fold split of data) 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).

5. Discussion and conclusions

Some conclusions from the results are as follows. The best models were Gaussian Process model and those BNN with Student's t-distribution noise model without ARD prior distribution. The early-stopped committee MLP ESC models were not able to sufficiently control the complexity of the model comparing with Bayesian neural networks models. 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.

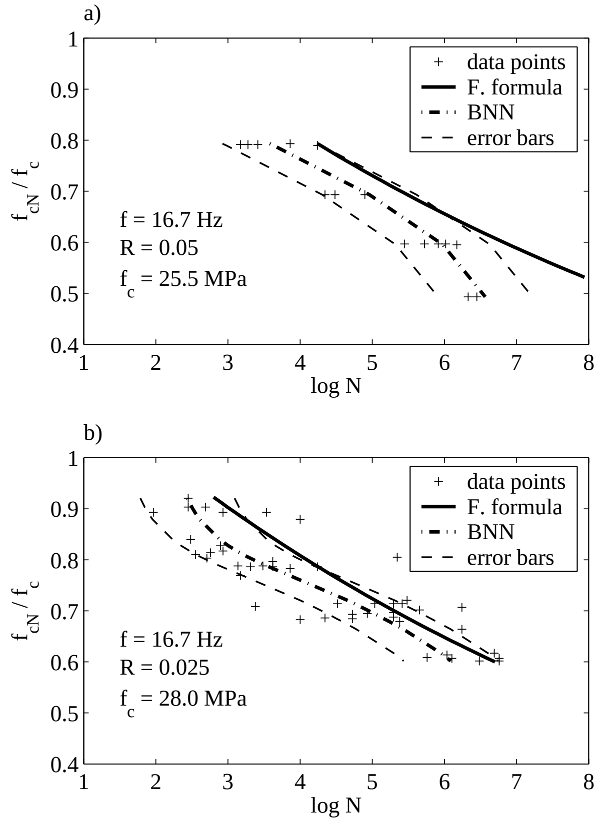

Fig. 1. Comparison of concrete fatigue durability prediction of relations $f_c - \log N$ for tests performed by: a) Gray et al. (1961), b) Antrim and Mc Laughlin (1959) taken from [3].

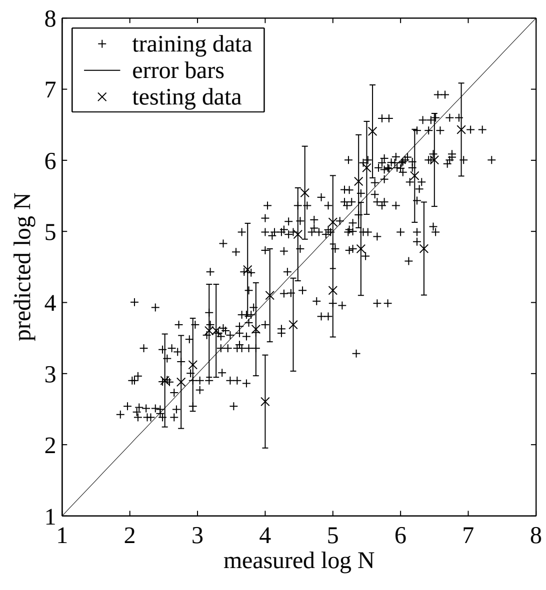

Fig. 2. Concrete fatigue durability values measured vs predicted by Bayesian neural network with 1σ error bars (for 22 testing patterns only)

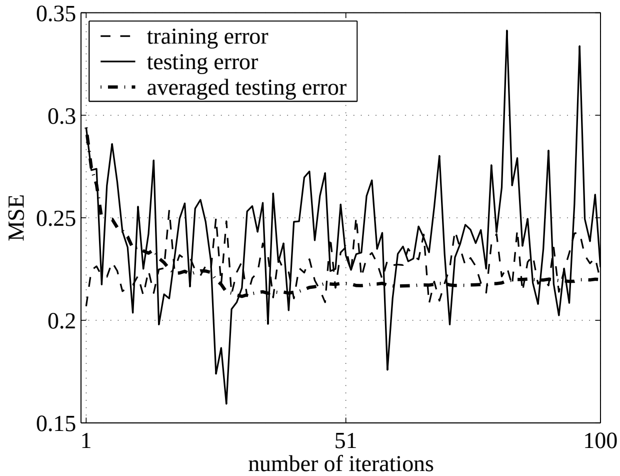

Fig. 3. Mean square error on the training and testing sets for networks sampled in the last 100 iterations (one of the runs)

The results presented in the present paper demonstrate the feasibility of using a Bayesian neural network model to predict concrete fatigue durability. The uncertainty in the model's predictions is due largely to noise in the training data. The input data were noisy and sparse. 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.

[2] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.

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

[4] 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.

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

[6] J. Lampinen and A. Vehtari. Bayesian approach for neural networks – review and case studies. *Neural Networks*, 14(3):7–24, April 2001. (Invited article).

[7] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[8] I. T. Nabney. *Netlab: Algorithms for Pattern Recognition*. Springer-Verlag, London, 2002.

[9] R. M. Neal. Bayesian training of backpropagation networks by the hybrid Monte Carlo method. Technical Report CRG-TR-92-1, 1992.

[10] 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, editors, *Neural Networks and Soft Computing*, Advances in Soft Computing. Physica-Verlag (Springer), 2002.

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

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