<u>Enclosure 13 to TN E-28034</u>

H. Fujii and K. Ichikawa: "Estimation of Weld Properties by Bayesian Neural Network," Welding in the World, vol. 70 (2001), No.3, p.335-339 (associated with RAI 2-9)

Welding International 2001 15 (12) 935-939
Selected from Journal of Japan Welding Society 2001 70 (3) 335-339; Reference JL/01/4/335; Translation 2858

Estimation of weld properties by Bayesian neural network

H FUJII
Joining and Welding Research Institute, Osaka University

K ICHIKAWA
Steel Research Laboratories, Nippon Steel Corporation

Introduction

This study was initiated from a conception such that it would be beneficial to construct a visionary model and computational tools which could inform us of the weld properties, such as, the strength, hardness and toughness etc. immediately after input of the welding conditions.

The weld properties are affected in a complex manner by an extremely large number of factors, such as the various material properties and welding conditions, so it is very difficult to construct an analytical model to facilitate the above. The weld properties depend not only upon variables which can be determined as test conditions, such as welding speed and specimen material composition but also upon many potential variables which are difficult to be quantified or of which we are not aware.

Neural networks have the characteristics such that various nonlinear relations are combined and extremely complex functions can be constructed, so this is an ideal technique to deal with complex phenomena.[1] In principle, any set of complex functions can be approximated within a certain range of errors. Thus, to date a number of attempts have been made to use this technique to control, for example, domestic appliances and automatic welding.[1-4]

The authors perceived this flexibility of functions and conceived the construction of a system which could evaluate the weld properties using the neural networks; subsequently, in June 1998, the Ad hoc Research Committee for the 'Estimation of Weld Properties by Bayesian Neural Networks' was inaugurated.[5]

Information on neural networks to date indicates that data dispersion and errors due to curve fitting had been dealt with ambiguously, and so it has not been possible to apply this technique to problems where data dispersion is crucial, such as the prediction of weld properties. Accordingly, in order to solve this problem for this study, a system which can be applied to predict weld properties has been constructed by the use of new type neural networks with the concept of Bayesian estimation[6] which was added to conventional neural networks. It became feasible by the use of this technique to predict error bars to the predicted results. This research committee was initially expected to last for 1 year but an extension of a further year was later agreed; in this way, sufficient results were thought to be achieved so that numerous papers and verbal presentations were put forward.[7-14]

An outline of neural networks was given in the previous report;[5] Consequently, in this report, Bayesian techniques which could not be introduced previously due to the limited space is introduced as a central feature.

Neural network structure

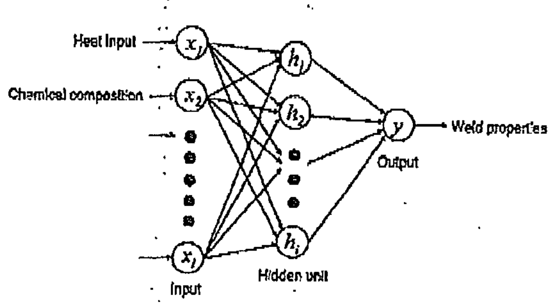
As introduced in the previous report[5], the structure of the neural network employed in this study is shown in Fig. 1. With the input of experimental conditions, such as the heat input and the chemical composition of specimen material, the output, that is the weld properties, can be estimated. Hidden units can be employed between the input and the output and complex functions can be expressed by controlling the values of these. The relationship between the respective input x_j and the hidden unit h_i of the order i can be generally expressed by a nonlinear function as in the following:

$$h_i = \tanh\left(\sum_j w_{ij}^{(1)} x_j + \theta_i^{(1)}\right) \qquad [1]$$

The relationship between the hidden unit h_i and the output y is linear as in the following:

$$y = \sum_i w_i^{(2)} h_i + \theta^{(2)} \qquad [2]$$

The weight coefficient w and the threshold value θ for these functions are optimised using database, as described later. As can be seen, with the neural network it is possible to construct various complex functions including, in general, nonlinear relationships, with the combination of nonlinear functions.



1 *Structure of the neural network.*

The weight coefficient w and the threshold value θ are determined so that the following energy function becomes a minimum.[15,16]

$$M(w) = \beta E_D + \alpha E_W \quad [3]$$

Where the parameter vector w is made up of the weight coefficient w and the threshold value θ. In addition, α and β, as described later, are parameters which control the complexity of the model. The energy function consists of the error function E_D and the regularizer E_W.

The error function is the total of the square of the difference between the estimated value and the experimental value as indicated in the following.

$$E_D(w) = \frac{1}{2}\sum_m (y(x^{(m)};w) - t^{(m)})^2 \quad [4]$$

Where $\{x^{(m)}, t^{(m)}\}$ is the data set and $x^{(m)}$ indicates the input variable and $t^{(m)}$ shows the experimental data, in other words, the target. m is the label of the combination of the data and the target. When the model is well consistent with the data, in other words, $y(x^{(m)}; w)$ is close to $t^{(m)}$, the error function E_D becomes minimum.

E_W is the total of the square of the weight coefficient.[17]

$$E_w(w) = \frac{1}{2}\sum_i w_i^2 \quad [5]$$

These regularizers serve to make $y(x^{(m)};w)$ become a smooth function of x. This term facilitates to lessen w and also lessens the tendency for the model to become overfitting to the data dispersion. Overfitting means that, as a result of an attempt of fitting to all the dispersed data, the result is the construction of an over-complex function; reference should be made to this aspect which was presented in the previous report.

Method to determine the Bayesian coefficients α, β and the number of hidden units

When β is excessively large, the degree of freedom of a function increases and overfitting is likely to occur. Conversely, when α increases, a function becomes too smooth and does not fit to the data. As can be seen, α and β are very important parameters; however, from results of testing the neural network to date, these values had been established in a nearly appropriate manner.

With the use of the Bayesian estimation, the statistical implications can be given for the determination of α and β as in the following. First, under the conditions where a certain data D would occur, the conditional probability p(w|D) where a certain combination w is likely to occur for the weight coefficient w and the threshold value θ can be expressed, according to the Bayesian equation as in the following.[6]

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \quad [6]$$

Here, in order to determine the most probable weight coefficient w and the threshold value θ, p(w|D) should be made to become maximum. According to equation (6), there is the following relationship

$$p(w|D) \propto p(D|w)p(w) \quad [7]$$

so, the probability is obtained on the assumption that there is a dispersion in accordance with the normal distribution for each of the right side. The normal distribution can be expressed by equation (8), supposing the average is m and the standard deviation is σ.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-m)^2}{2\sigma^2} \quad [8]$$

Thus, the dispersion of the data in the case when the weight coefficient w and the threshold value θ are expressed by a certain value w, can be expressed by the following equation.

$$p(D|w) = \prod_{m=1}^{N} p(t^{(m)}|x^{(m)},w) = \frac{1}{Z_D}\exp\left(-\frac{1}{2\sigma_v^2}\sum_{m=1}^{N}(y(x^{(m)},w) - t^{(m-)})^2\right) \quad [9]$$

At this time $x^{(m)}$ is the input variable and $t^{(m)}$ is the experimental data, in other words, a target; Z_D indicates the normalised constant and σ_v is the data dispersion. Here, after substituting equation (4) for equation (9), equation (10) is obtained.

$$p(D|w) = \frac{1}{Z_D}\exp\left(-\frac{1}{\sigma_v^2}E_D\right) \quad [10]$$

On the other hand, there is a dispersion even with p(w) and its probability can be expressed by the following equation.

$$p(w) = \frac{1}{Z_W}\exp\left(-\frac{1}{2\sigma_W^2}\|w\|^2\right) = \frac{1}{Z_W}\exp\left(-\frac{1}{\sigma_w^2}E_W\right) \quad [11]$$

Here, equation [5] was substituted. Zw is the normalised constant and σ_w is the dispersion from the real value of w.

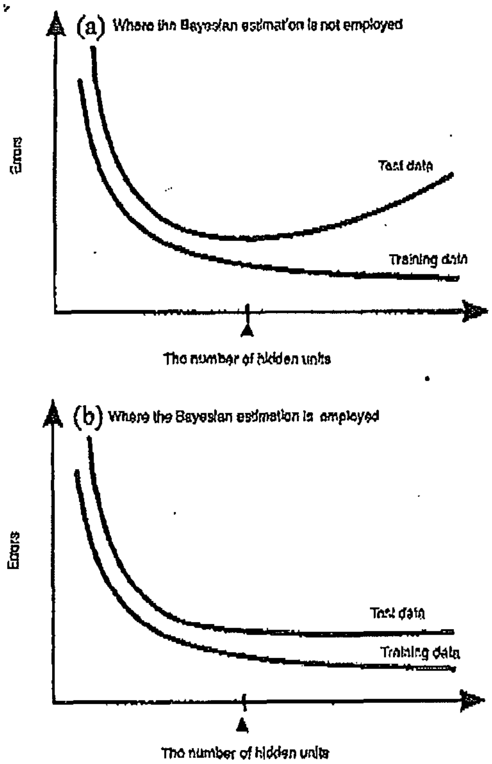
When equations (10) and (11) were substituted for equation (7), the following equation was obtained.

$$p(w|D) \propto p(D|w)p(w) = \frac{1}{Z_D Z_W} \exp\left\{-\left(\frac{1}{\sigma_v^2}E_D + \frac{1}{\sigma_w^2}E_W\right)\right\} \quad [12]$$

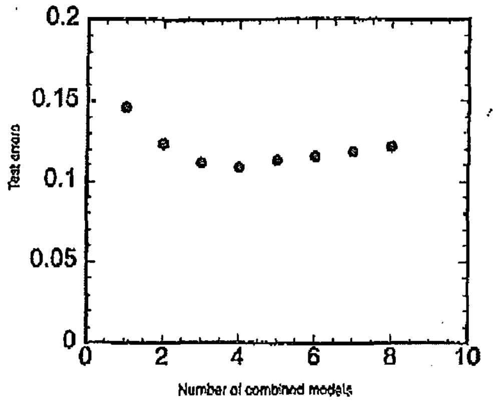
Consequently, in order to make p(w|D) a maximum, $\left(\frac{1}{\sigma_v^2}E_D + \frac{1}{\sigma_w^2}E_W\right)$ should be made minimum. In equation [3], by comparison with the energy function $M(w) = \beta E_D + \alpha E_W$ which is to be made minimum, it is evident that α and β bear the following statistical implication.

$$\beta = \frac{1}{\sigma_v^2},\ \alpha = \frac{1}{\sigma_W^2} \quad [13]$$

Thus, from this, accurate training becomes possible.



2 Variations in errors due to the number of hidden units: (a) For a common neural network, (b) For a neural network where the Bayesian estimation is employed.



3 Relationship between the number of models and errors in committee model.

As can be seen, control parameters α and β control the complexity of the model but the complexity of the model can be controlled by varying the number of hidden units. The latter method is employed in the general neural network. The more complex is the input and output relationship, the more hidden units are required; however, as described in the previous report[3], when the number of hidden units are increased excessively in order to raise the accuracy of estimation, overfitting will occur and the accuracy will conversely deteriorate.

The following method was considered: in order to prevent overfitting to a minimum, half of the data is randomly selected and training is applied to the neural network using the selected data only, the remaining half is employed as test data in order to examine the optimisation of the model. The magnitude of error in the training data set and test data set varies, as indicated in Fig. 2, with increasing numbers of hidden units. The difference between the estimated value and the experimental value of the training data simply decreases with increasing number of hidden units, but the difference in the test data at first falls and then increases. A large error when there are less hidden units indicates the impossibility of fitting due to the functions being too simple and an increased test error when there are many hidden units indicates that functions are overfitting. Accordingly, the best model is selected for the case where the test error takes the minimum value and this model is employed for prediction of weld properties.

The increase due to overfitting is very small with the use of the Bayesian estimation. In principle, overfitting will not occur even though the number of hidden units is infinite, if the data in use complies with the Gaussian distribution and the Bayesian model is fully optimised.[1,8] In other words, the variation is given as seen in Fig 2 (b).

Committee model

In order to improve further the estimation accuracy, a committee model, which consists of a combination of multiple models, was employed in this study.

In neural networks, although little different from the optimised model in respect of errors, models which have entirely different structures, for example, many models with varied numbers of hidden units can be formed. Thus, by a combination of these, defects of each single model can be mutually compensated for and the estimation accuracy can be improved.

First of all, using error functions, these models are to be ranked as the optimised model, the second optimised model, the third optimised model and so on. Then, using equation [14], the models are successively complemented from the optimised model and the number of models to be complemented is to be increased until the error becomes minimal; the model in which the error reaches minimum is to be selected as the committee model which is then to be used for the estimation.

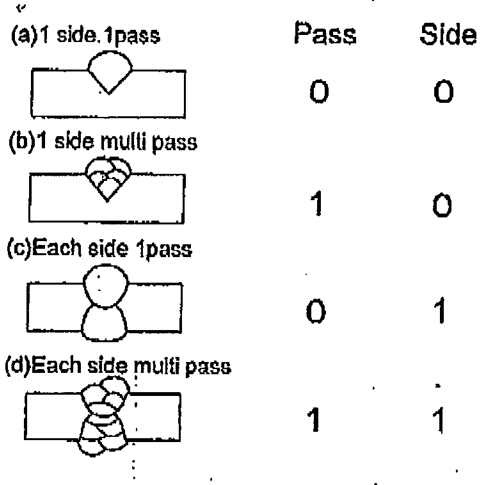
$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad [14]$$

The error bar of the committee model is calculated using equation [15].

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 + \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \qquad [15]$$

where N: the number of models, y_i,σ_i: the estimated values and the error bars of individual models.

Figure 3 shows the results, as an example, of the construction of the committee model of tensile strength. It is evident from the figure that the models up to the fourth



4 Method for the numerical expression of the welding processes.

have decreasing errors by combination with the committe model but after complementing the models of the fifth and thereafter, the errors conversely start to increase. Subsequently, under the circumstances, a combination of the optimised model to the fourth optimised model is selected as the committee model. The predicted error on this occasion proves to be as much as 25.4 % less compared with the case of the optimised model on its own. Accordingly, the estimation accuracy markedly improves by the use of the committee model.

Weld properties

When an appropriate neural network is constructed, the properties can be estimated under optional conditions. The systems were developed in this study for the steel weld metal strength, fracture appearance transition temperature (FATT) and hardness. The weld properties can be varied by variations to the welding conditions in addition to the material properties, so they are affected by an extremely large number of parameters. However, the more complicated the neural networks, the more they were able to demonstrate their efficacy; the weld properties were a very appropriate subject of neural network application. Using this system each individual property can be predicted, together with the error bar, following input of numerial values as the input parameters, as indicated in Tables 1–3 of the previous report[5].

Furthermore, in the neural network, every input data should be expressed numerically. Subsequently, for example, the welding process also influences the extent of the weld metal reheat and greatly influences the mechanical properties, so it is necessary to define this numerically as input variables so that it can be readily processed for engineering applications. This was hardly taken into consideration for the conventional regression analysis; however, in this study, 4 welding processes, as shown in Fig. 4, were classified by 2 parameters which are the number of welding passes and the number of welding sides. As can be seen, for examples where there is no significance in the numerical value itself, it is necessary to employ values which were represented by 2 values, in other words, either 0 or 1 in order to prevent prediction accuracy from degrading. In addition, on this occassion, it is desirable to keep the number of parameters to the minimum.

In a neural network training itself requires time, but once appropriately structured, prediction can be completed in a short time under optional conditions. In addition, using this neural network, the significance of individual factors can be estimated by the use of σW as indicated in equations [11] and [13].[15,16] When the value of σW is large, the input factor relating to this causes a relatively significant change to the output. Such evaluation is regarded to be very useful in optimisation of welding conditions and composition.

The publication of data is as described in the previous report[5] but typical prediction results are shown in references 10–14. Here, in this paper, prediction was carried out with the range of prediction 1 σ, in other words, with a confidence limit of 67 %. The magnitude of the error bar varies greatly according to the input conditions. In this manner, for examples where the computer has no confidence in its prediction, the error bar is enlarged and it has the effect to stimulate the user and draw the user's attention to the result. Under the circumstances, the cause is either the data dispersion under the conditions of use or insufficent data.

Conclusions

As described so far, neural networks have a capacity to reconstruct the data base and it is believed that it will be possible to hand information down to the next generation of this feature by making full use of the accumulated data base and know-how.

By incorporating the Bayesian estimation into conventional neural networks, the prediction of weld properties was attempted and adequate results in welding engineering were obtained. In addition, as one of the characteristics of this prediction, it was was identified that unreliableness of estimation can be displayed by the magnitude of the error bar. According to this system, the magnitude of the error bar depends upon the input conditions (test conditions) at the time; for example, where the data dispersion is large and reliability is low, the error bar is displayed as large and the computer itself is equipped with a function which can display the reliability of its prediction.

The prediction of this error bar substantially extended the application scope of neural networks and allowed the possibility of application to the reconstruction of data bases of various properties. Neural networks are very powerful tools and can reconstruct various data bases, not necessarily just those of weld properties. Anyone who has an interest is invited to contact us.

References

1 Baba et al: 'Fundamental knowledge and application of neural networks'. Kyoritsu Shuppan 1994 6.
2 Yakawa: 'Neural networks'. Baifukan 1992.

3 Eguchi K et al: *Sci Tech Weld Join* 4 (6) 327.
4 Suga Y et al: *ISIJ Int.*39 (10) 1075.
5 Fujii H and Ichikawa K: *J Jpn Weld Soc* 2000 69 (3) 44.
6 Lindley D V: 'Introduction to probability and statistics from a Bayesian view-point'. Cambridge University Press 1965.
7 Fujii H et al: *Proc Mater Adv Power Eng* 1998 1401 Liege.
8 Fujii H et al: J Jpn Metal Soc 1999 63 (7) 905.
9 Yurioka et al: IIW Doc IX-1930-99 1999.
10 Fujii H et al: Abstracts of lectures given to the 1999 Autumn National Conference of the Japan Welding Society 1999 65 516.
11 Fujii H et al: Abstracts of lectures given to the 1999 Autumn National Conference of the Japan Welding Society 1999 65 518.
12 Fujii H et al: *Welding Technique* 2000 48 (1) 82.
13 Fujii H et al: Abst. Int. Conf. Processing and Manufacturing of Adv. Mater. 2000 4–8 Dec. Las Vegas USA 88.
14 Fujii H et al: *J Mater Process Tech* in press.
15 MacKay D J C: *Neural Computation* 1992 4 415.
16 Fujii H et al: *ISIJ Int* 1996 36 (11) 1373.
17 MacKay D J C: *Neural Computation* 1992 4 448–472.
18 Neal R M: Radford Thesis 'Bayesian learning for neural networks' Dept. of Computer Science, Univ. of Toronto 1995 122.