USE OF BP NEURAL NETWORK TO PREDICT HYDROGEN CONTENT IN COAL

Ju L.C.(Xian Jiaotong University, China), Tade M.O. (Curtin University, Australia) and Zhu J.N. (Curtin University, Australia)

Abstract: Neural network is a popular method for predicting unknown process variables from measured process data. Many learning algorithms have been proposed in the literature to improve model prediction. In this paper, we introduce the concept of sample study risk in neural network (NN) to improve the prediction of hydrogen content in coal using Back Propagation (BP) NN. Targeting the problem of training convergence quality impaired by the interfering information of some samples in BP NN, the validity of the concept of sample study in NN and the feasibility of analysing the chemical element in coal using NN are discussed.

Keywords: neural network, BP neural network, sample study risk, hydrogen content in coal.

1. INTRODUCTION

The hydrogen content in coal is of great significance for indicating the quality of coal. It is especially important for a combustion process. However, direct determination of hydrogen content is not an easy task. Ultimate analysis, though can give individual elemental content including the hydrogen, the method is nevertheless rather complicated and in most cases it has to be conducted in an external laboratory. Therefore it is difficult to apply ultimate analysis frequently during coal fired power generation. On the contrary, proximate coal analysis is a simple process and is usually available everywhere. This analysis gives the most basic information on coal properties, including contents of moisture, fixed carbon, ash, volatile matter and lower heating value. Unfortunately, the information regarding elemental content is not provided. Nonetheless, it is speculated that the proximate analysis and ultimate analysis have some connections and these connections will be exploited using Artificial Neural Network (ANN) in this paper. Consequently, the ANN prediction model of hydrogen content in coal, the model training and the model checking by proximate analysis of coal, using the Back Propagation (BP) arithmetic of Neural Network (NN) are discussed.

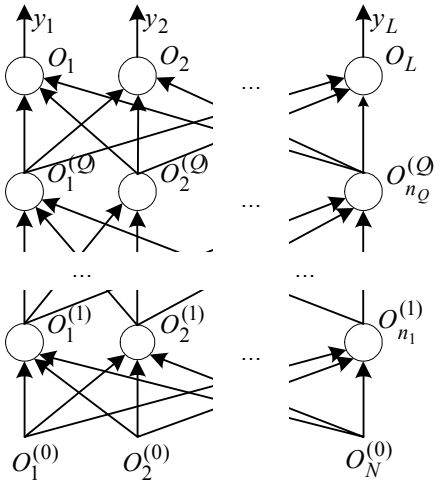

Fig.1 Structure of BP Neural Network

2. BP NEURAL NETWORK MODEL

The Back Propagation Neural Network (BP NN) structure adopted is shown in Figure 1. For the given volume of training samples: $(x_{p1}, x_{p2}, \cdots x_{pN}) \rightarrow (t_{p1}, t_{p2}, \cdots t_{pL}), p = 1,2\cdots$, (where P is the sample number, N, L are the input/output vector dimensions respectively), the average sum of square error between the results of the NN operation and the output of the training samples is:

$$E = \frac{1}{P} \sum_{p=1}^{P} E_p \qquad (1)$$

$$E_p = \frac{1}{2}\sum_{l=1}^{L}\left(t_{pl} - y_{pl}\right)^2 \qquad (2)$$

where, t_{pl} is the expected output of the number l output cell in the number p sample; y_{pl} is the results of NN operation of the number l output cell in the number p sample.

The process of training study by the network includes the forward calculation in the inner network and the backward error propagation calculation. These calculations minimize the NN error by adjusting the inner connections of the NN. For the case of a single hidden layer BP NN and multi-layer BP NN connections between the implicit layer and the output layer is adjusted according to the rule of delta [Haykin, (2001)]. The connection between the hidden layers or between the output layer and hidden layer of the multi-layer BP NN is manipulated by the BP arithmetic [Sun et al, (1997), Mills, et al (1996)] (an advanced arithmetic) suggested by Rumelhart and others [Haykin, (2001)].

2.1 Formula for calculating network node

As shown in Figure 1, the forward calculation of the BP inner network is given by:

(1) The output of nodes in the input layer

For N dimensions, the input vector is

$$O_i^{(0)} = \left(x_i - \left(\max\left(x_i\right) + \min\left(x_i\right)\right)/2\right) / \left(\max\left(x_i\right) - \min\left(x_i\right)\right) + 0.5 \qquad (3)$$

Where, max (x_i) and min (x_i) is the maximum and minimum value of the ith sample among all the samples.

(2) The input/output of nodes in hidden layers

For implicit layers the input of the number q is:

$$I_i^{(q)} = \sum_{j=1}^{n_{q-1}} w_{ij}^{(q)} O_j^{(q-1)} - \theta_i^{(q)} \qquad \left(i = 1,2,\cdots,n_q\right) \qquad (4)$$

Where, $w_{ij}^{(q)}$ is the connecting weight from the number j node of the number (q – 1) layer to the number i node of the number q layer; $\theta_i^{(q)}$ is the threshold value of the number i node in the number q layer; $O_j^{(q-1)}$ is the output of the number j node in the number (q – 1) layer; n_{q-1} is the number of nodes in the number (q – 1) layer.

The output of the nodes in the number q layer is:

$$O_i^{(q)} = f\left(I_i^{(q)}\right) \qquad \left(i = 1,2,\cdots,n_q\right) \qquad (5)$$

Where, f is the Sigmoid function: $f(x) = \left[1 + \exp(-x)\right]^{-1}$; $i = 1,2,\cdots, n_q$, n_q is the number of nodes in the number q layer.

(3) The input/output of the output layer

If the output vector has L dimensions, then the input of the nodes in the output layer is:

$$I_i = \sum_{j=1}^{n_Q} w_{ij} O_j^{(Q)} + \theta_i \qquad \left(i = 1,2,\cdots,L\right) \qquad (6)$$

Where, w_{ij} is the connecting intention from the number j node of the number q hidden layer to the number i node of the output layer; θ_i is the threshold value of the output layer; $O_j^{(Q)}$ is the output of the number j node in the number Q layer; n_Q is the number of nodes in the number Q layer.

The output of the nodes in the output layer is:

$$O_i = f\left(I_i\right) = \left[1 + \exp\left(-I_i\right)\right]^{-1} \qquad \left(i = 1,2,\cdots.L\right) \qquad (7)$$

2.2 The iterative computation of connection weights

(1) The connection weight of the nodes between the output and the hidden layers is given by

$$w_{ij}(k+1) = w_{ij}(k) + \alpha\left[(1-\eta)D_{ij}(k) + \eta D_{ij}(k-1)\right] \qquad (8)$$

$$D_{ij}(k) = \delta_i O_j^Q \qquad (9)$$

$$\delta_i = \left(t_i - O_i\right)O_i\left(1 - O_i^Q\right) \qquad (10)$$

Where, α is the study rate (step length), and $O < \alpha < 1$; η is the momentum factor, and $O \le \eta < 1$; $D_{ij}(k-1)$ is the negative gradient at the (k – 1) time step; $i = 1,2,\ldots, L$, $j = 1,2,\ldots, n_Q$.

(2) The connection weight between the hidden layers is

$$w_{ij}^{(q)}(k+1) = w_{ij}^{(q)}(k) + \alpha\left[(1-\eta)D_{ij}^{(q)}(k) + \eta D_{ij}^{(q)}(k-1)\right] \quad (11)$$

$$D_{ij}^{(q)} = \delta_i^{(q)} O_j^{(q)} \quad (12)$$

$$\delta_i^{(q)} = \left[\sum_{k=1}^{n_{q+1}} w_{ki}^{(q+1)} \delta_k^{(q+1)}\right] \bullet O_i^{(q)}\left[1 - O_i^{(q-1)}\right] \quad (13)$$

Where, q=0,1,…,Q, when q = 0, its input layer, and $n_0 = N$; $w_{ij}^{(q)}$ is the connecting weight from the number j node of the number (q - 1) hidden layer to the number i node of the number q layer; k is the iteration time.

To minimize the training error of the network (defined by equation (2)) and speed up the convergence of the training, the method of variable study rate is adopted. When $\Delta E_n = E_n - E_{n-1} < 0$, we choose a bigger study rate, and do the converse for the opposite situation. This can be expressed as follows:

$$\begin{cases} \text{if} \quad \Delta E_n < 0, \text{ then } \alpha = \alpha a, \quad a > 1 \\ \text{else } \alpha = \alpha b, \quad 0 < b < 1 \end{cases}$$

Where, "a" and "b" are constants; α is a variable study rate. This training method can automatically adjust the study rate according to the training error and solve the problem of poor convergence. One only needs to choose the study rate once by experience. The values of "a", "b" and α can be acquired by computer calculation according to the above principle. The optimised value of "a", "b" and α that was obtained from the literature [Yao et al, (1996)] is applied in this paper, α = 0.87, a = 1.05, b = 0.68.

For the general case the correct prediction model can be obtained using the method of variable step length. However, due to the complexity of coal property, even though the momentum item and the method of variable step length study are introduced, the quality of convergence in the training is not perfect. This is because the characteristic parameters of some coal sample are greatly different from the other samples. In other words, some samples can have greater incoherence than others. To describe such incoherence, the concept of study risk in the NN studies is adopted in this work.

3. THE RISK IN NN SAMPLE STUDY

During the course of BP network's online study, suppose the present network after P group samples $\mathbf{x}_p = [x_{p1}, x_{p2}, \ldots, x_{pn}]^T$, $\mathbf{d}_p = [d_{p1}, d_{p2}, \ldots, d_{pL}]^T$, (p = 1.2,….,P; **x** is the input and **d** is the output.) training, the matrix of connecting weights is $\mathbf{W}_p$ and the network's output is: $\mathbf{y}_p = [y_{pl}, y_{p2}. \ldots, y_{pL}]$. When the new sample (P + 1) is received, after one training, the matrix of connecting weights becomes $\mathbf{W}_{p+1}$ and the corresponding connecting weights of the previous P groups of samples' output is: $\mathbf{y}_{p+1} = [y_{(p+1)1}, y_{(p+1)2}, \ldots, y_{(p+1)L}]$, then the associated risk to the network due to the new study sample can be expressed as:

$$\sigma = \frac{1}{2}\sum_{p=1}^{P}\sum_{l=1}^{L}\left(y_{(p+1)l} - d_{pl}\right)^2 \Big/ \left[\frac{1}{2}\sum_{p=1}^{P}\sum_{l=1}^{L}\left(y_{pl} + d_{pl}\right)^2\right] \quad (14)$$

σ indicates the degree of consistency of the new sample P + 1 and the previous P groups of samples. Small σ shows that new sample's mapping relations is approaching the fixed BP NN's and the training is easier to reach convergence after receiving new sample with lower risk. On the contrary, a big σ shows that the new sample includes different information from the old ones and the training will not easily achieve convergence. Accordingly, the training is of a higher risk. Therefore, σ is called the study risk parameter for this sample.

Since the first P groups of samples are not based on a ready network, a hypothesis is necessary to confirm the study risk. Taking a network model, which has taken certain time to train with a comparatively stable error, as the basic network, then the study risk parameter for number p(p ≤ P) group samples' can be expressed as:

$$\sigma_i = \frac{1}{2}\sum_{p=1}^{P}\sum_{l=1}^{L}\left(y_{(p+1)l} - d_{pl}\right)^2 \Big/ \left[\frac{1}{2}\sum_{p=1}^{P}\sum_{l=1}^{L}\left(y_{pl} - d_{pl}\right)^2\right] \quad p \neq i, i \leq P \quad (15)$$

where $y_{(p+1)l.}$ ($l \neq i$) is the network's output after the basic network's first training with the number i sample, y_{pl} is the output of the basic network.

When the training error is relatively stable during the training process, the network can be said to have acquired certain optimisation. If further training can result in greater changes to other samples, it is certain that the connecting weights of the network have changed significantly relative to the basic network model. The consistency can be described by σ_i as shown in equation (15).

The study risk parameter of the sample indicates the degree of the dissimilarity of the sample with the other ones. The risk comes from two sources, namely: one is the samples contain significant interrupting information; the other is that the samples contain some new information that is not indicated by the other samples. It is necessary to control a large risk since it directly affects the convergence of BP NN. On the condition that the samples do not change, a large risk indicates that the sample contains significant interrupting information. In order to reduce the risk these interrupting information should be removed. However, this measure is on the other hand likely to sacrifice the opportunity of learning new object properties. Therefore, both the convergence of the network and the learning of new object properties have to be taken into account. If convergence is ensured, use of a large risk parameter is recommended. Controlling the study risk is sometimes very effective to screen out the influence of some large interrupting information and improve the convergence of the network. The following section illustrates how to use the concept of study risk on the prediction of hydrogen content in coal.

4. THE PREDICTION OF HYDROGEN CONTENT IN COAL USING NN

Since coal qualities usually vary from seam to seam, it is more difficult and impossible to predict the hydrogen content for any coal with a single NN model without compromising the precision of the prediction. In order to overcome this difficulty, we separate the coals into several categories according to the differences in ratios of volatile matter, ratios of ash content, ratios of dry air moisture and the ratios of lower heating values. NN is then applied to predict the hydrogen content of each type of the coal respectively.

(1) Selection of the structural parameters of the BP NN

The structural parameters of BP NN include layers and nodes. The numbers of input nodes and output nodes are decided based on the complexity of the process. The key problem is the number of hidden layers and nodes required. Unfortunately, there is no theory available so far to select this number. The ability of the network to approach non-linear function will increase with the enhancement of layers and nodes, but the amount of calculation increases rapidly. Therefore, the structural parameters of BP are obtained on the basis of large number of experiments. In this paper, we adopt four layers structure: one input layer, two hidden layers and one output layer. There are four nodes in the input layers, eight and four nodes in the hidden layers respectively, and one node in the output layer. The nodes of the input layers are the ratios of volatile matter, ratios of ash content, ratios of dry air moisture content and the ratios of lower heating values, respectively. According to the principle of section 2, the coefficient of study risk is confirmed to be 1.1 through repeated computation.

(2) Prediction results of hydrogen content in coal based on BP NN

We used the data of 904 groups of coals to gain the training of design and testing of the model. Considering the study risk of NN, we excluded 61 groups of sample data due to their larger study risk parameter and adopted 720 groups of data as the training sample. After the training of the parameters to the allowable range of error, the validity of the model is verified by the remaining 123 groups of data. The prediction result is shown in Figure 2 for the 123 samples with the model after training. In the figure, the abscissa is arranged in the increasing order of coal's ratio of volatilisation. The prediction error is shown in Figure 3. The biggest error is 8.7%, which is in the allowable range.

For a good model, the square difference of the estimated error should be equal or less than 10^{-2}. We can use the following formula as the square difference of the error [Bakirtzis et al, (1996) and Hou et al, (1996)].

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^{n} e^2(k) \qquad (16)$$

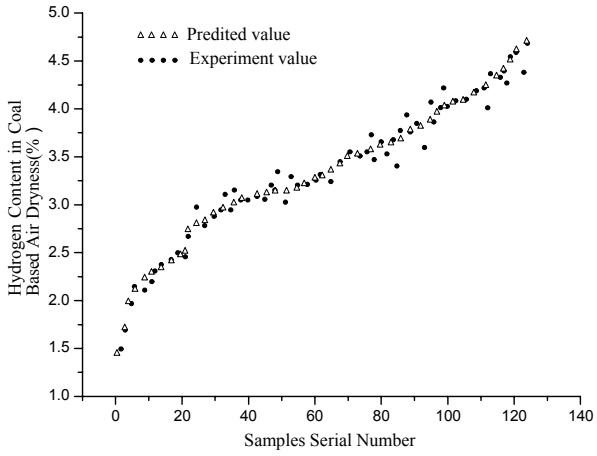

Fig.2 Prediction Results of Hydrogen Content in Coal based on Neural Network

where e(k) is the difference between the predicted value and the measured value. The square difference of the estimated error is shown in Table 1. From Table 1, we can see that the standard square difference of the estimated error for this model is less than 10^{-2}, which demonstrate that the NN can better predict the amount of hydrogen in the coal.

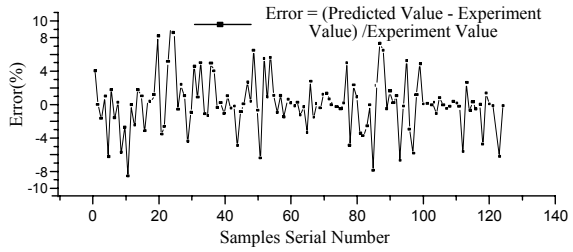

Fig.3 Prediction Errors of Hydrogen Content in Coal Based on Neural Network

Table 1: Validity Verification of Predicted Model

n	$\bar{y}$	squared difference	standardised squared difference
123	3.27	0.017	$0.0016\ \bar{y}^2$

In the above results only 123 samples are not training samples in all the samples, and hence they are the ones used to establish the validity and reliability of the model. The standard square difference of the estimated error for the model is also less than 10^{-2} when less samples were used for the checking (i.e. n < 123 in Table 1). Furthermore, the standardised squared difference of the estimated error for the model also improves substantially for n > 123 at the expense of increasing computational load.

5. CONCLUSIONS

NN has great potential for modelling non-linear systems and uncertain processes. In this paper, we proposed the concept of sample study risk for NN, and applied it to the prediction of the hydrogen content in the coal. The control of sample study risk can accelerate the convergence of training for the NN and accurate prediction of the hydrogen content in coal also demonstrates that the control of study risk is valid for the convergence of the trained NN.

6. REFERENCES

Hagkin, S. (2001). *Neural networks, a comprehensive foundation*. 2nd Ed. Tsinghua University Express, China.

Sun, Z., Zhang, Z and Deng, Z., (1997). *The theory and technology of aptitude control.* Tsinghua University Express and The Science and Technology Express. Guangxi.

Yao, R., Sun, G., and Tang. C., (1996). *An improved BP method in the NN for multi-layers.* The transactions of South-East University. Vol 26, No. 4. pgs 126-130.

Mills, P.M., Zomaya, A.Y. and Tadé, M.O. (1996). *Neuro-Adaptive Process Control – A Practical Approach.* John Wiley & Sons, New York.

Bakirtzis, A.G., Petridis, V., and Kiartzis, S. (1996). *A neural network short term load predicting model for the Greek power system. IEEE Trans on Power Systems.* Vol. 11, No. 2. pgs 858-863.

Hou, Z., Wu, J., and Zhang, Q., (1996). *Several improved means for the short term burthen in the electric power system.* The Automation of the Electric System. Vol. 20, No. 7. pgs 27-31.