Prediction Model of Total Organic Carbon Content on Hydrocarbon Source Rocks in Coal Measures Established by BP Neural Network Based on Logging Parameters

Pan Wang, Wenfeng Du, Mingxing Liang, Hongwei Wang, Wenxi Wei
State Key Laboratory of Coal Resources and Safe Mining
China University of Mining & Technology (Beijing)
Beijing, China
Wangpan328@yahoo.com

***Abstract*—The total organic carbon content (*TOC*) is an important parameter for source rocks evaluation in coal measures. Total organic carbon content determined from logging parameters using back propagation neural network technique, which provide a new method for hydrocarbon source rock evaluation. Use the Turpan Basin Xishanyao formation as the research object. The five logs which consist of volume gamma logging (*GR*), acoustic logging (*AC*), density logging (*DEN*), resistivity logging (*RT*) and compensation neutron logging (*CNL*) were selected optimally based on the correlation analysis of the total organic carbon content measured data and well logging parameters as the input vector of BP neural network, and the total organic carbon content was selected as the output vector of BP neural network. Then the BP neural network model was established and applied to predict total organic carbon content for Xishanyao formation of B1 well in the Turpan Basin, with a competitive analysis of the prediction errors. The error between prediction values and measured values is small, and the majority of the relative errors are less than 8%. The results show that the BP neural network model based on logging with optimal parameters has a very strong generalization ability, and can approximate the nonlinear relationship between total organic carbon content and logging parameters of coal measure source rocks with high accuracy.**

Keywords*—coal measures; logging parameters; hydrocarbon source rocks; BP neural network; total organic carbon content

I. Introduction

The total organic carbon content is one of the significant parameters in the evaluation of hydrocarbon source rocks. The conventional method to obtain total organic carbon content is to test and analyze the finite sample and take the mean value to represent the organic carbon content of the whole set formation. But the heterogeneity of hydrocarbon source rocks makes the conventional method very undesirable because the organic carbon content of the rock strata obtained by taking the average value cannot reflect the true situation of the content of organic carbon in the source rock[1]. In addition, the limited sample analysis data is difficult to obtain a continuous total organic carbon content data. Considering the characteristics of high resolution and vertical continuous of logging technology, we can establish the relationship between well logging information and organic carbon content to calculate the organic carbon content of the hydrocarbon source rocks[2]. It is very complicate to express the nonlinear relationship between hydrocarbon source rocks which has many difference lithology and well logging parameters. The artificial neural network has a strong nonlinear approximation ability, which can describe the relationship between the input and output[3]. Therefore, the core content of total organic carbon content and logging parameters are analyzed to optimize logging parameters. Establishment the relationship between total organic carbon content and logging parameters by BP artificial neural network to obtain total organic carbon content prediction model[4]. Using the model to predict the data of the organic carbon data gap can obtain the longitudinal continuously and accurate total organic carbon content data.

II. Theory and Method

Coal measures source rocks contain a lot of organic petrology, total organic carbon content difference cause the difference of log information. Based on sampling data, the correlation coefficient between total organic carbon content and log parameters is defined as equation (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(\mathrm{y}_i - \overline{\mathrm{y}})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(\mathrm{y}_i - \overline{\mathrm{y}})^2}} \tag{1}$$

where r denotes the correlation coefficient, $\overline{x}$ 、 $\overline{\mathrm{y}}$ are the average value, x_i 、 y_i denote the *i*th sample point corresponding to the core logging observations. By selecting the larger relevant coefficient as the learning sample parameters, we could predict the total organic carbon content through the most sensitive parameters.

Utilization of district 7 logs and total organic carbon content data to do correlation analysis and cross-correlation of each log analysis, the correlation coefficient obtained (Table 1). According to correlation analysis, preferably a gamma logging (*GR*), resistivity logging (*RT*), acoustic logging (*AC*), density logging (*DEN*) and compensation neutron logging (*CNL*) logging data as BP neural network input data

Artificial neural network is a network composed of many artificial neural networks. In essence, the network implements a mapping function from the input to the output[5]. It has been proved that it has the ability to realize any complex nonlinear mapping theory in mathematics. The artificial neural network can be divided into mutual connection type and hierarchical network according to the connection way of network neurons. At present, the application of hierarchical network is more extensive. The BP neural network is based on the difference between the actual output and the expected output, which is a multilayer feed forward network where the error correction is carried out from the back to front step by step for the weights and thresholds of each layer of the network[6]. This method regards the unknown system as a black box. First of all, the input and output data of the system are used to train the BP neural network to make sure that the essence of the unknown function is the minimum value of the error function. Through repeated training of a number of known samples, the training stops when the error reaches the desired value[7]. At this time, the weights of each layer and the threshold value of each layer neuron are obtained as knowledge. So the training process is over. Then we can use the trained BP neural network to predict the output of the system and the whole calculation process is as shown in figure 1.

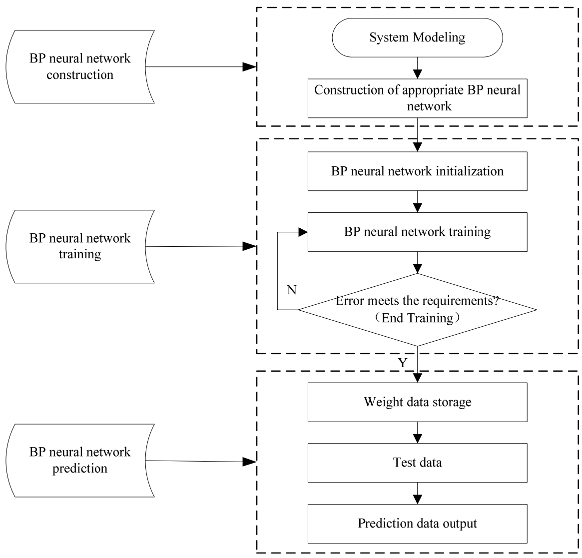

Fig.1 The computational flow chart of BP neural network

During BP neural network training, a set of parameters which consists of the network layer number, the hidden layer node number, the number of iterations, the initial weight as well as the learning rate and the expected error, etc. needs to be considered and selected correctly. Rohet-Nielson has proved that a three layer neural network which includes the input layer, the hidden layer and the output layer can approximate the continuous function with any precision. This network model has only one hidden layer and the hidden layer can be composed of one or more layers of nodes. The network is not connected with the same layer nodes, but is connected with the next layer nodes[8]. When it is working, input data gets into input layer nodes and then in turn pass over the nodes in the hidden layer and finally transmitted to the output layer nodes. The output data of each node will be treated as the input data of the next node. Sigmoid function is used as the activation function of neurons in the network. Training network using sample measured data and the objective function is defined as the error between actual output and desired output[9]. If the prediction error does not meet the requirements of error, error feedback back to the new training by adjusting the weights and thresholds, the training was stopped until the error meets the requirements.

According to the actual situation of the research, three layers of BP networks with five importing units and one exporting unit were set. Before the training, the numerical model of the learning data was normalized with the equation (2).

$$x_n^* = \frac{x_n - x_{\min} + c}{x_{\max} - x_{\min} + c} \tag{2}$$

where x_n is measured data, x_{min}, x_{max} denote the minimum and maximum values of the measured data respectively, x_n^* is the data after normalization, c is the correction parameter defined as a constant.

By the preferred logging parameters, the input layer has 5 nodes and the output parameter is *TOC* content. The output layer has one node, and through cut-and-trial the nodes in hidden layer are set to 9. So it is obviously that BP neural network structure is 2-9-1. We select the hyperbolic tangent function as activation function, and use dynamic learning rate to make training error convergent rapidly by iteration. The model of the BP neural network is defined as equation (3).

$$TOC = \sum_{j=1}^{m} \left[v_{jt} \tanh\left(\sum_{i=1}^{n} w_{ij} x_i + a_j\right) \right] + b_t \tag{3}$$

where xi is the logging parameter of 5 network input layer, including gamma ray logging (*GR*), acoustic logging (*AC*), resistivity logging (*RT*), density logging (*DEN*) and compensation neutron logging (*CNL*) logs respectively, w_{ij}, a_j are the weight coefficient and threshold from the network input layer to the hidden layer, v_{jt}, b_t are the weight coefficient and threshold from the hidden layer to the output layer, n=5 is the number of feature vectors in the input layer, m=9 denotes the nodes of hidden layer, t=1 is the number of output layer.

The weight coefficient and threshold value could be obtained shown as Table 2 when BP neural network model training stops.

III. EXAMPLE

In the study, 67 samples' total organic carbon content and logging data were used to establish the neural network model, random select 54 samples as learning samples, and the left 13 samples drop out of neural network training, they used for testing the network. Checking the calculation model's generalization ability. The Fig.2 is the congruent map of predict output data and expect output data, we can see that the well trained neural network model has very good predict results. Making relative error graph to visualize the forecast error. (Fig.3) The cartoon reveals that neural network model predict results' relative error less than 8%, once again shows the predict ability of neural network model. Finally, the established neural network model is applied to B1 well's total organic carbon content prediction, obtain the exact, continues total organic carbon content data (Fig.4).

Fig 2 The measured values and predicted outputs values of model

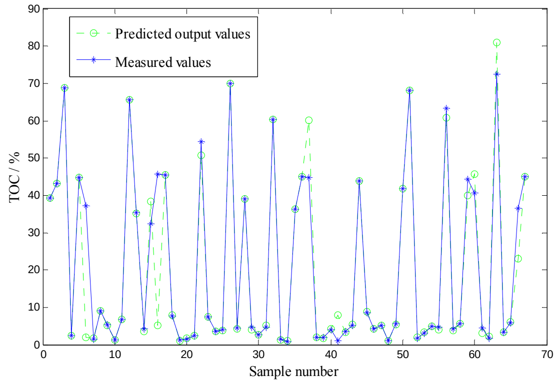

Table 1 Correlation Coefficient between TOC Content and Log Parameters

parameters	TOC	RT	SP	CAL	GR	DEN	CNL	AC
TOC	1							
RT	0.344	1						
SP	0.225	0.178	1					
CAL	0.313	0.269	0.183	1				
GR	0.487	0.376	0.307	0.148	1			
DEN	0.778	0.03	0.338	0.577	0.443	1		
CNL	0.711	0.033	0.515	0.513	0.286	0.868	1	
AC	0.746	0.113	0.388	0.565	0.301	0.618	0.505	1

Table 2 Weight Coefficient and Threshold of BP Neural Network Model

weight coefficient	weight coefficient from the network input layer to the hidden layer w_{ij}					threshold a_j
	1	2	3	4	5	
1	0.4518	-0.6791	-0.4707	1.7781	1.5060	-2.5865
2	-101462	-1.0352	-1.5972	-0.5987	1.4086	1.9816
3	0.9849	0.9267	2.8940	-1.5727	-2.0062	-0.8266
4	-0.0537	-1.9631	1.2271	0.4117	1.0209	1.6150
5	1.4938	-0.8123	-0.5468	1.7722	-0.3976	0.0696
6	-0.5445	1.1621	-0.4960	1.1161	-2.1946	-0.3625
7	1.0842	-0.5840	-2.9231	-1.5431	2.0031	1.4640
8	0.1709	2.3650	1.2640	1.7220	2.1463	-4.1159
9	-0.3066	-0.9154	0.9339	1.1200	-1.2302	-2.3968

weight coefficient and threshold from the hidden layer to the output layer v_{jt}									Threshold b_t
1	2	3	4	5	6	7	8	9	
-0.6704	-0.2252	1.2640	0.2807	1.1056	0.3058	1.9932	1.6963	-0.9423	0.6718

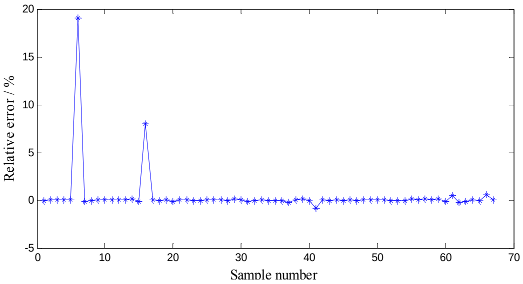

Fig 3 The relative errors between the measured values and predicted outputs values

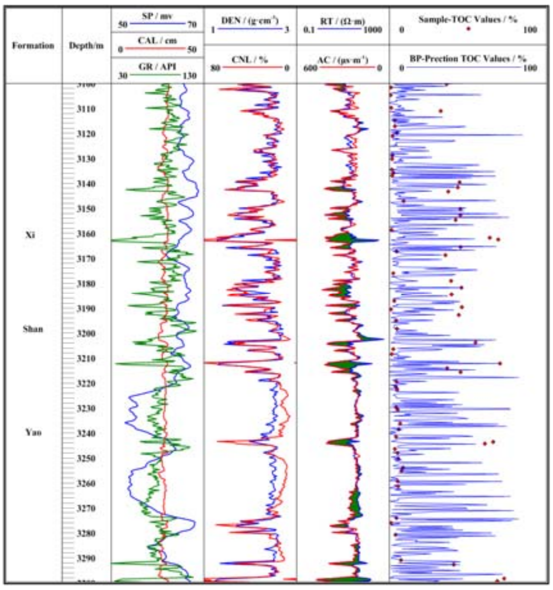

Fig 4 The application of prediction model in well B1

IV. CONCLUSION

Organic carbon content has a certain relationship with logging parameters, through the correlation analysis reveal that the organic carbon content of coal measure strata has a very good relationship with *GR*, *RT*, *AC*, *DEN*, and *CNL* logs.

The non-linear relationship between total organic carbon content and logging parameters is very complex. Using single logging parameter to establish organic carbon predict model has a very low precision. So we must consider all the logging parameters to establish the organic carbon predict model.

Using BP neural network's approximation capability to complex nonlinear function, establish the BP neural network that base on the optimization logging parameters to predict the total organic carbon content of coal measure strata, obtain very good results. The relative error basic below 8%.

ACKNOWLEDGMENT

We are grateful for financial supports from National Natural Science Funds Coal Joint Project (U1261203), National Key Scientific Instrument and Equipment Development Program (2012YQ030126), Shanxi Natural Science Funds Project (2013012001).

REFERENCES

[1] B. Alizadeh, S. Najjari, and A.K. Llkhchi, "Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran," Computer&Geosciences, vol. 45, pp. 261-269, August 2012.

[2] M.R. Kamali, and A.A.Mirshady, "Total organic carbon content determined from well logs using ΔLogR and Neuro Fuzzy techniques," Journal of Petroleum Science and Engineering, vol. 45, pp. 141-148, August 2004.

[3] H.A. Bakhtiar, A. Telmadarreie, M. Shayesteh, M.H. Heidari, H. Talebi, and Z. Shirband, "Estimating total organic carbon content and source rock evaluation, applying ΔlogR and neural network methods: Ahwaz and Marun oilfields, SW of Iran," Petroleum Science and Technology, vol. 29, pp. 1691-1704, July 2011.

[4] D.M. Jarvie, R.J. Hill, T.E. Ruble, and R.M. Pollastro, "Unconventional shale-gas system: The Mississippian Barnett Shale of north-central texas as one model for thermogenic shale-gas assessment," AAPG Bulletin, vol. 91, pp. 475-499, April 2007.

[5] D.J.K. Ross, and R.M. Bustin, "Characterizing the shale gas resource potential of Devonian-Mississippian strata in the Western Canada sedimentary basin: Application of an integrated formation evaluation," AAPG Bulletin, vol. 92, pp. 87-125, January 2008.

[6] Q.R. passey, S. Creaney, J.B. Kulla, F.J. Moretti, and J.D. Stroud, "A practical model for organic richness from porosity and resistivity logs," AAPG Bulletin, vol. 74, pp. 1777-1794, December 1990.

[7] E. Ghasemi-Nejad, M.J. Head, and M. Naderi, "Palynology and petroleum potential of the Kazhdumi Formation (Cretaceous: Albian–Cenomanian) in the South Pars field, northern Persian Gulf," Marine and Petroleum Geology, vol. 26, pp. 805-816, June 2009.

[8] H. Rahimpour-Bonab, "A procedure for appraisal of a hydrocarbon reservoir continuity and quantification of its heterogeneity," Journal of Petroleum Science and Engineering, vol. 58, pp. 1-12, August 2007.

[9] A. Bhatt, and H.B. Helle, "Committee neural networks for porosity and permeability prediction from well logs," Geophysical Prospecting, vol. 50, pp. 645-660, November 2002.