doi: 10.3969/j.issn.1673-9736.2016.01.02
Article ID: 1673-9736(2016)01-0006-07

Estimation of reservoir porosity using probabilistic neural network and seismic attributes

HOU Qiang, ZHU Jianwei* and LIN Bo

College of Earth Sciences, Jilin University, Changchun 13006, China

Abstract: Porosity is one of the most important properties of oil and gas reservoirs. The porosity data that come from well log are only available at well points. It is necessary to use other method to estimate reservoir porosity. Seismic data contain abundant lithological information. Because there are inherent correlations between reservoir property and seismic data, it is possible to estimate reservoir porosity by using seismic data and attributes. Probabilistic neural network is a powerful tool to extract mathematical relation between two data sets. It has been used to extract the mathematical relation between porosity and seismic attributes. Firstly, a seismic impedance volume is calculated by seismic inversion. Secondly, several appropriate seismic attributes are extracted by using multi-regression analysis. Then a probabilistic neural network model is trained to obtain a mathematical relation between porosity and seismic attributes. Finally, this trained probabilistic neural network model is implemented to calculate a porosity data volume. This methodology could be utilized to find advantageous areas at the early stage of exploration. It is also helpful for the establishment of a reservoir model at the stage of reservoir development.

Key words: porosity; seismic attributes; probabilistic neural network

1 Introduction

Intelligent system is a powerful tool to extract quantitative formulation between two data sets. And it has started to be applied to the petroleum industry in recent years (Karimpouli *et al.*, 2010; Tahmasebi & Hezarkhani, 2012; Chithra *et al.*, 2013; Asoodeh & Bagheripour, 2014). There are inherent correlations between reservoir properties and seismic attributes (Yao & Journel, 2000; Huang & Wang, 2013; Iturrarán-Viveros & Parra, 2014; Li & Liu, 2015). Therefore, it is possible to estimate reservoir porosities by using seismic data and attributes. Previous studies have shown that it is feasible to estimate reservoir porosity by using statistical methods and intelligent systems (Leite & Vidal, 2011; Iturrarán-Viveros, 2012; Na'imi *et al.*, 2014).

Probabilistic neural network (PNN) is a neoteric neural network model based on statistical theory (Miguez, 2010). The training time of PNN is shorter and the accuracy is higher than traditional multilayer forward network. It is particularly suitable for nonlinear multi attributes analysis. For this case, PNN has excellent performance on unseen data. In this study, the propounded methodology is implemented to estimate the porosity of sandstone reservoir prosperously.

Received 28 October 2015, accepted 5 December 2015
* Corresponding author (zhujw1095@sina.com)

2 Probabilistic Neural Network

PNN is a variant of radial basis function networks and approximate Bayesian statistical methods. The process of PNN is similar to human behavior (Parzen, 1962). Probabilistic neural network is an alternative type neural network (Specht, 1990). It is based on Parzen's probabilistic density function estimator. PNN is a four-layer feed-forward network, consisting of an input layer, a pattern layer, a summation layer and an output layer (Muniz *et al.*, 2010).

Probabilistic neural network is actually a mathematical interpolation method, but it has the structure of neural networks. It has better interpolation function than multilayer feed forward neural network. The training data sample of PNN includes a number of training sample sets, and each sample corresponds to the seismic sample in the analysis window of each well.

Suppose that there is a data set of n samples, each sample consists of m seismic attributes and one reservoir parameter. Probabilistic neural network assumes that each output well log value could be expressed as a linear combination of input logging data value (Hampson *et al.*, 2001). The new sample after the attribute combination is expressed as:

$$x = \{A_{1j}, \quad A_{2j} \cdots\cdots A_{mj}\} \qquad (1)$$

The new predicted logging values can be expressed as:

$$\overline{L}(x) = \frac{\sum_{i=1}^{n} L_i \exp[-D(x, x_i)]}{\sum_{i=1}^{n} \exp[-D(x, x_i)]} \qquad (2)$$

where:

$$D(x, x_i) = \sum_{j=1}^{m} \left[\frac{(x_j - x_{ij})}{\sigma_j}\right]^2 \qquad (3)$$

The unknown quantity $D(x, x_i)$ is the "distance" between input point and each training sample point. This distance is measured by seismic attributes in multidimensional space and it is expressed by the indefinite quantity σj. Eq. (1) and Eq. (2) represent the application of a probabilistic neural network. The training process includes identifying the optimal smoothing parameter set. The goal of the determination on these parameters is to enable the validation error minimization.

Defining the kth target point validation result as follows:

$$\overline{L_k}(x_k) = \frac{\sum_{i \neq k} L_i \exp[-D(x_k, x_i)]}{\sum_{i \neq k} \exp[-D(x_k, x_i)]} \qquad (4)$$

When the sample points are not in the training data, it is the k-th target sample prediction value. Therefore, if the sample values are known, we can calculate the prediction error of sample points. Repeat this process for each training sample set, we can define the total prediction error of training data as:

$$E(\sigma_1, \sigma_2 \ldots\ldots \sigma_m) = \sum_{i=1}^{n} (L_i - \overline{L_i})^2 \qquad (5)$$

The prediction error is dependent on the choice of parameter σ_j. This unknown quantity realizes the minimization through nonlinear conjugating gradient algorithms. Validation error, the average error of all excluded wells, is the measure of a possible prediction error in the process of seismic attributes transformation. The trained Probabilistic neural network has the characteristics of validation minimum error.

The PNN is not an iterative learning process. So it has faster training speed than other artificial neural network architecture (Muniz *et al.*, 2010). This is a feature of the Bayesian technique (Mantzaris *et al.*, 2011).

3 Methodology

The data sets used in this study belong to 8 wells (i. e. W1 to W8) and post-stack 3D seismic data in the Songliao Basin, Northeast China. The target stratum is the first member of the Cretaceous Nenjiang Formation that is one of the principal reservoirs in this area. In this study, the main contents include seismic impedance inversion, attributes extraction, training and application of PNN model. The flow chart is shown in Fig. 1.

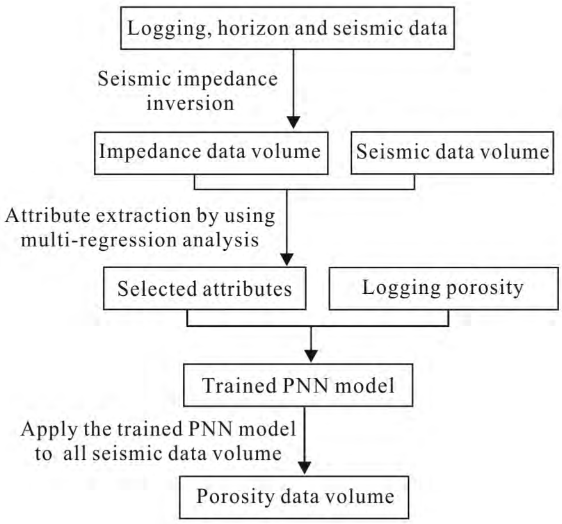

Fig. 1 Flow chart of this study

3.1 Seismic attributes selection by using multi-regression analysis

Multi-regression analysis is a mathematical method which is used to analyze the relationship between one dependent variable and several independent variables (Hampson *et al.*, 2001). The basic principle is that although there is no strict, deterministic functional relation between dependent variables and independent variables can try to find the most appropriate mathematical formula to express this relation.

Multi-regression analysis can be used to solve the following problems:

(1) Determine if there is a correlation between certain variables. If it exists, find a suitable mathematical expression of them.

(2) According to one or several variable values, predict the value of another variable, and calculate the forecast accuracy.

(3) Factor analysis. For example, in the common effects of many variables on a variable, find out the most important factors, the secondary important factors, and the relationship between these factors.

In the multi-regression analysis method, prediction error of N attributes is always less than or equal to N-1 attributes. Adding attributes means to use higher polynomial to fit the curve. We can calculate the prediction error of each polynomial. This prediction error is equal to the root mean square error between real values and predicted values. With the increase of the polynomial order, the prediction error decreases. But when we use over high order polynomial to fit curves, the existing data may fit well, but the interpolation or extrapolation over the boundary would be fitting badly. This problem is called "over-trained".

In this study, the data would be divided into a training data set and a validation data set. The training data set is used to determine the correlation coefficient, and the validation data set is used to compute the validation error. If a high order polynomial fit the training data set well, but fit the validation data set badly. It means that the order for polynomial is too high (Fig. 2).

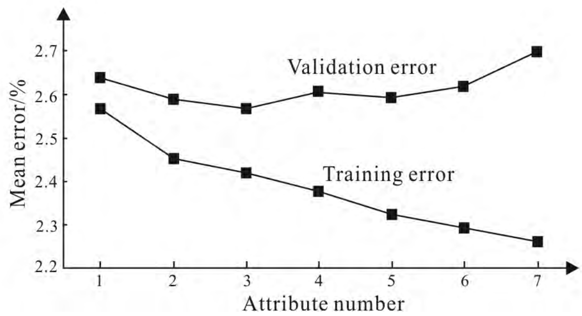

Fig. 2 Selection of attributes number

In this section, multi-regression analysis method is utilized to find the most suitable seismic attributes. As illustrated with Table 1, the training error gradually reduces with the increasing number of attributes, but when the number of attributes increases to four, validation error will rise. So, the best set of seismic attributes should contain three attributes that are the first three attributes in Table 2. The first three attributes are inverted impedance, average frequency and filter 35/40-45/50.

The most significant seismic attribute is inverted impedance. Those attributes yield useful information about the lateral changes in lithology and porosity (Chopra & Marfurt, 2005). Furthermore, the training error for them is less than 3% that shows the exactness of the results.

It should be pointed out that PNN is a kind of nonlinear method, so the aforementioned attributes can be used as input for porosity prediction by PNN (Kadkhodaie-Ilkhchi *et al.*, 2009).

Table 1 Result of multi-regression analysis for porosity estimation

Final attribute	Training Error/%	Validation Error/%
Inverted impedance	2.5677	2.6371
Average frequency	2.4526	2.5881
Filter 35/40-45/50	2.4191	2.5672
Integrated absolute amplitude	2.3775	2.6071
Filter 25/30-35/40	2.3238	2.5925
Instantaneous Phase	2.2927	2.6179
Second derivative instantaneous amplitude	2.2628	2.6973

3.2 Porosity estimation using PNN

The main purpose of this section is to establish an optimum PNN model. The inputs to this model are three selected attributes in the previous section. In order to highlight the advantages of probabilistic neural network in porosity estimations, another four algorithms have been used. Another four algorithms are single attribute analysis, multi-regression analysis, multi-layer feed forward network (MLFN) and radial basis function (RBF). The training and validation results are shown in Table 2. According to the results, PNN algorithm gives less training and validation error. As seen from Table 3, the correlation coefficient of training result could reach 0.915, which is considered as a high correlation coefficient. It is higher than multi-regression analysis method (the correlation coefficient of multi-regression analysis is 0.844) and other methods. According to the numerical validation results, PNN method of porosity estimations is more accurate than others in this case. In the final of this section, the analysis for creating an optimum PNN model was done (Table 2 and Fig. 3).

Table 2 Training and validation results of neural networks

method	Training result		Validation result	
	correlation	Error/%	correlation	Error/%
single attribute analysis	0.822	2.568	0.811	2.637
multi-regression analysis	0.844	2.420	0.822	2.567
PNN	0.915	1.806	0.881	2.227
MLFN	0.864	2.280	0.848	2.676
RBF	0.886	2.086	0.772	2.963

4 Results and discussion

We have demonstrated the application of the probabilistic neural network to reservoir porosity estimations by seismic attributes. Two mathematical tools have been used: multi-regression analysis and PNN method.

In the section of seismic impedance inversion, a qualified inverted impedance data volume has been calculated. In the section of seismic attributes selection, multi-regression analysis has been used to find appropriate seismic attributes (Table 1). Those seismic attributes come from 3D seismic data volume and inverted impedance data volume. The optimal model

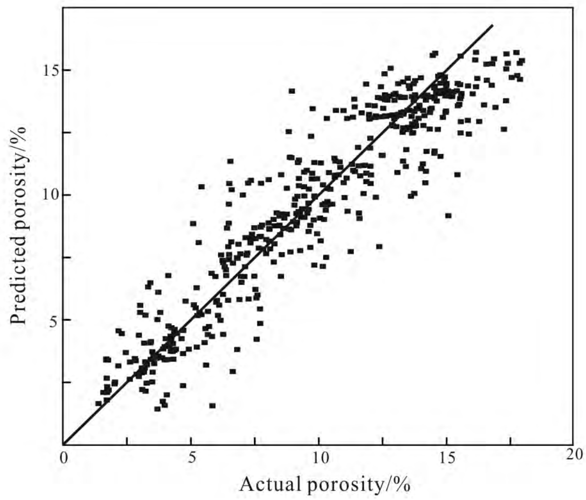

Fig. 3　Cross plot of predicted porosity versus actual porosity

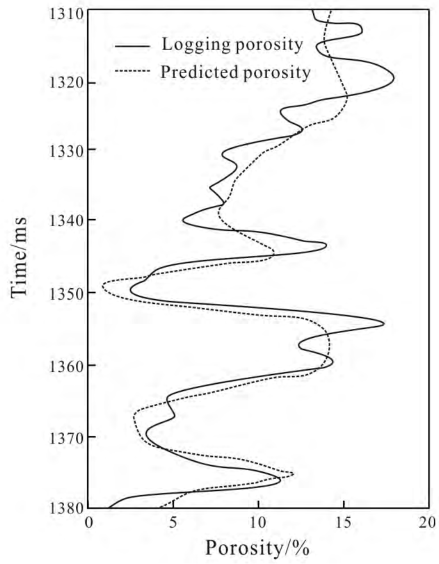

Fig. 4　Validation result of W5

is built up by PNN with proper trends and minimization of error.

We have demonstrated this methodology on a set of 8 wells log data. The correlation coefficient of training data set could reach 0.915, which is considered as a high correlation coefficient (Fig. 3).

The well W5 is not used in training. It is used to validate the result of porosity estimations. The correlation coefficient of validation result could reach 0.881, which means that this methodology is reliable. The estimated porosity of W5 is displayed in Fig. 4: solid line is the original porosity and dotted line is the esti-

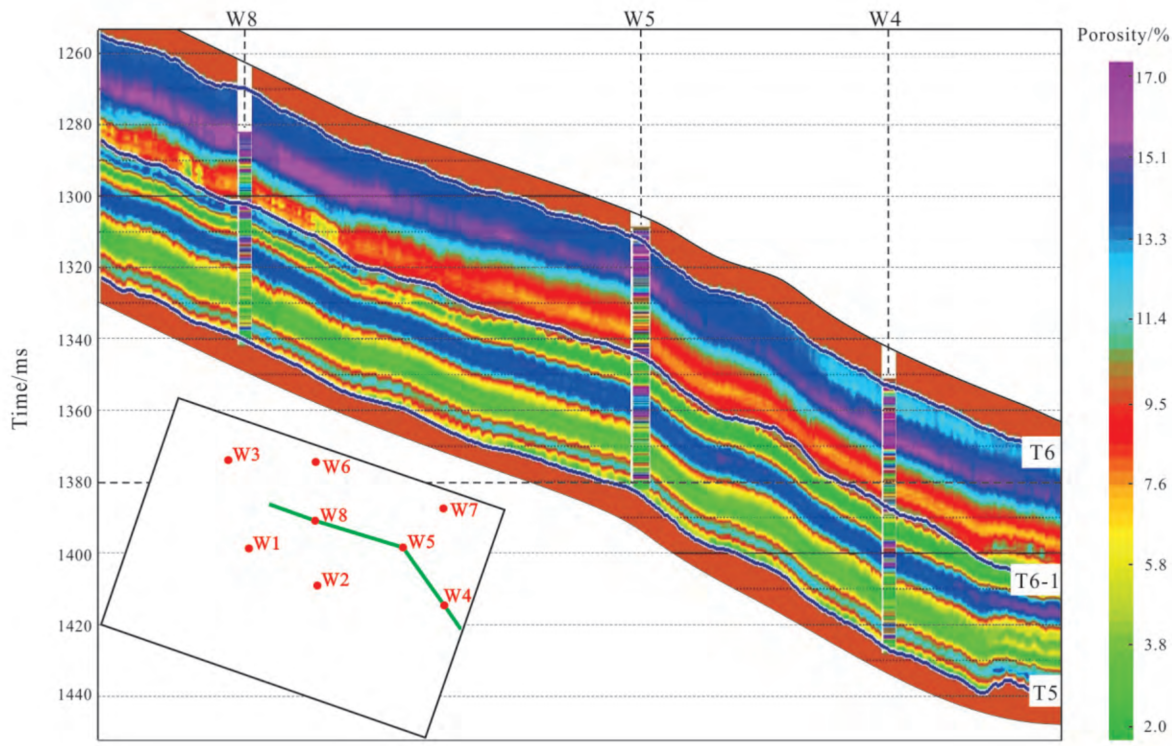

Fig. 5　Arbitrary line from porosity data volume

mated porosity. The correlation coefficient and the error are 0.881 and 2.227%, respectively.

After the establishment of an optimum PNN model for porosity estimations, we apply this model to all seismic data volume. Then, porosity data volume could be calculated (Figs. 5, 6). In Fig. 6, an ancient river could be seen in the rectangle with higher porosity than elsewhere in the region. This is consistent with the law of geology, which shows, from one aspect, that the probabilistic neural network is a reliable tool for porosity estimations. This method is an effective way to create an acceptable porosity data volume.

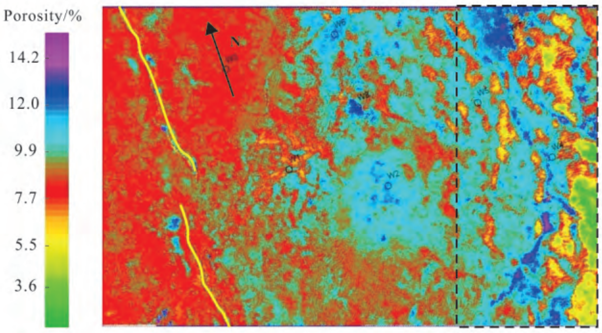

Fig. 6 Porosity map of T6-1 from porosity data volume

5 Conclusions

We have demonstrated that the estimation of reservoir porosity from seismic attributes and inversion impedance using PNN method. In this study, two mathematical tools have been used: multi-regression analysis and PNN method. On attributes selection stage of this study, three attributes have been selected. At the porosity estimation stage, a PNN model has been established and trained. The training and validation correlation coefficient between predicted porosity and actual porosity could reach 0.915 and 0.881, respectively. The profile of estimated porosity shows that porosity variation in the vertical direction is approximately increasing from the bottom to the top and can be verified at well locations.

The results indicate that PNN is a reliable method of porosity estimations. And it has obvious advantages in estimation accuracy compared with conventional methods such as multi-regression analysis and multi-layer feed forward network. The proposed methodology can be used to estimate porosity from seismic data. This methodology could reduce drilling risks and improve the success rate of exploration at the initial stage of reservoir exploration. And it also could provide an acceptable porosity data volume which could be utilized to build reservoir geological models at the stage of reservoir development.

References

Asoodeh M, Bagheripour P. 2014. ACE stimulated neural network for shear wave velocity determination from well logs. *Journal of Applied Geophysics*, **107**(4): 102-107.

Chithra C N, Song K Y, Gupta M M, *et al*. 2013. An innovative neural forecast of cumulative oil production from a petroleum reservoir employing higher-order neural networks (HONNs). *Journal of Petroleum Science and Engineering*, **106**: 18-33.

Chopra S, Marfurt K J. 2005. Seismic attributes—a historical perspective. *Geophysics*, **70**(5): 3SO-28SO.

Hampson D P, Schuelke J S, Quirein J A. 2001. Use of multiattribute transforms to predict log properties from seismic data. *Geophysics*, **66**(1): 220-236.

Huang W, Wang Z G, Dong J H, *et al*. 2013. Application of seismic data for forecasting hydrocarbon source rocks in Shahezi Formation of Xujiaweizi fault depression. *Global Geology*, 2013, **32**(3): 571-578. (in Chinese with English abstract)

Iturrarán-Viveros U. 2012. Smooth regression to estimate effective porosity using seismic attributes. *Journal of Applied Geophysics*, **76**: 1-12.

Iturrarán-Viveros U, Parra J O. 2014. Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. *Journal of Applied Geophysics*, **107**: 45-54.

Kadkhodaie-Ilkhchi A, Rezaee M R, Rahimpour-Bonab H, *et al*. 2009. Petrophysical data prediction from seismic attributes using committee fuzzy inference system. *Computers & Geosciences*, **35**(12): 2314-2330.

Karimpouli S, Fathianpour N, Roohi J. 2010. A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). *Journal of Petroleum Science and Engineering*, **73**(3): 227-232.

Leite E P, Vidal A C. 2011. 3D porosity prediction from seismic inversion and neural networks. *Computers & Geosciences*, **37**(8): 1174–1180.

Li J L, Liu C, Guo Z Q, *et al*. 2015. Rock physics model and simulation of seismic AVO responses for a shale reservoir. *Global Geology*, **34**(2): 497–504. (in Chinese with English abstract)

Mantzaris D, Anastassopoulos G, Adamopoulos A. 2011. Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. *Neural Networks*, **24**(8): 831–835.

Miguez R, Georgiopoulos M, Kaylani A. 2010. G-PNN: A genetically engineered probabilistic neural network. *Nonlinear Analysis: Theory, Methods & Applications*, **73**(6): 1783–1791.

Muniz A M S, Liu H, Lyons K E, *et al*. 2010. Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait. *Journal of Biomechanics*, **43**(4): 720–726.

Na'imi S R, Shadizadeh S R, Riahi M A, *et al*. 2014. Estimation of reservoir porosity and water saturation based on seismic attributes using support vector regression approach. *Journal of Applied Geophysics*, **107**: 93–101.

Parzen E. 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics, **33**(3): 1065–1076.

Specht D F. 1990. Probabilistic neural networks. *Neural Networks*, **3**(1): 109–118.

Tahmasebi P, Hezarkhani A. 2012. A fast and independent architecture of artificial neural network for permeability prediction. *Journal of Petroleum Science and Engineering*, **86**: 118–126.

Yao T, Journel A G. 2000. Integrating seismic attribute maps and well logs for porosity modeling in a west Texas carbonate reservoir: addressing the scale and precision problem. *Journal of Petroleum Science and Engineering*, **28**(1): 65–79.