7th International Conference & Exposition on Petroleum Geophysics

"HYDERABAD 2008"

P - 80

Classification of Lithofacies Boundaries Using the KTB Borehole Data: A Bayesian Neural Network Modeling

[1]Saumen Maiti*, and [2]Ram Krishna Tiwari

[1]Indian Institute of Geomagnetism, Navi-Mumbai-410218, India

[2]National Geophysical Research Institute, Hyderabad-500007, India

Email: saumen_maiti2002@yahoo.co.in

Summary

A novel approach based on the concept of Bayesian neural network learning theory is developed and applied to German Continental Deep Drilling Program (KTB) well log signal for classification of lithofacies boundaries. We parameterized different combination of synthetic model to match with the published histogram lithofacies succession. A Multi Layer Perceptron (MLP) network model (with four layers e.g. input, output and two hidden layers) is found suitable for the present pattern recognition problem. Objective function is optimized following scaled conjugate gradient optimization technique and uncertainty about the relationship between input and output domain is appropriately taken care of by assuming Gaussian prior distribution of networks parameters. Posterior distribution of network parameter is estimated following the Bayesian probability theory. The stability and effectiveness of the new approach is also tested on "noisy" data mix with different level of colored noise. We show that Bayesian neural network approach outperformed conventional neural network approach in terms of uncertainty, over fitting and under fitting problem which remains the serious problem with conventional neural network approach. Our analyses demonstrate that the Bayesian Neural network based approach renders a robust mean for the classification of complex litho facies successions from the KTB borehole log signal and thus provide useful guide for understanding the crustal inhomogeneity and structural discontinuity in many other tectonically critical and complex regions.

Keywords: *Well Log, KTB bore hole, Neural Networks, MLP, Bayesian Neural Networks, Lithofacies.*

Introduction

Classification of lithology/lithofacies boundary from geophysical well log signal is a complex and nonlinear geophysical problem because of several factors, such as pore fluid, effective pressure, fluid saturation, pore shape, grain size etc nonlinearly affect the geophysical well log signals. Geophysical nonlinear problem is characterized by nonlinearity of physics and by statistical nature of the solutions, latter being partly due to the physics and partly due to noise in the geophysical measurements. Geoscientist has been engaged in classifying litho-facies units from the recorded well log signal using the conventional method like graphical cross-plotting and other statistical techniques (Rogers *et al.* 1992). In the graphical cross-plotting technique (Pickett 1963; Gassaway *et al.* 1989), two or more logs are cross-plotted to yield lithologies. Multivariate statistical methods such as principle component and cluster analyses (Wolff & Pelissier-Combescure 1982) and discriminant function analysis (Delfiner *et al.* 1987; Busch *et al.*1987) have invariably been used for the study of borehole data. These techniques are, however, semi-automated and require a large amount of data, which are costly and not easily available every time .Further the existing methods for well log analysis are also very tedious and time-consuming, particularly when dealing with large number of noisy and complex borehole data. Recently, Maiti & Tiwari (2005) have developed skilled algorithm based on Walsh transform technique for automatic detection of litholog boundary and, more recently, Maiti & Tiwari (2005) and Maiti *et al.* (2007) have developed neural network modeling and classification of lithology/lithofacies unit using conventional neural network and network stability in presence of correlated color noise. However, there are major limitations in the conventional neural network approach (Coulibaly *et al.* 2001a, Aires 2004). One of the main limitations is that the network is trained by maximizing a likelihood function of the parameters or equivalently minimizing an error function in order to obtain the best set of parameters starting with a initial random set of parameters. Sometimes a regularization term with an error function is used to prevent overfitting. In that method, a complex model can fit the training data well but it does not necessarily mean that it will provide smaller errors with respect to new data. This happens because of the uncertainty about the relationship between input and output mapped by the network during training. It is, therefore, imperative to search for a more appropriate nonlinear technique, which could evade these difficulties. Here, we develop a neural network program on Bayesian frame to classify litho-facies boundaries and apply the method to well log data from German Continental Deep Drilling Project (KTB).

The KTB Borehole Data

The German Continental Deep Drilling Project (KTB) explores a metamorphic complex in northeastern Bavaria, southern Germany. Lithologically, the continental crust at the drill site consists of three main facies units: paragneisses, metabasites and alternations of gneiss-amphibolites, with minor occurrence of marbles, calcsilicates, orthogeneisses, lamprophyres and diorites (Berckhemer *et al.* 1997; Emmermann and Lauterjung J. 1997;Peching *et al.* 1997;Leonardi and Kumpel 1998,1999) .Visual inspection indicate these KTB bore hole data show non linear signal characteristics (Figure-1). We, therefore, generate a suitable parameterized Bayesian neural network training model to match the published histogram model of three sets of recorded well log signal (viz, gamma ray, density and neutron porosity). (Table 1, see Maiti *et al.* 2007).

Table 1: Showing the significant limits to generate forward model for neural network training indicating that gamma ray intensity value most crucial factor to categorize litho-facies unit in metamorphic area

Litho-facies unit	Density (g/cc)	Neutron porosity (%)	Gamma Ray Intensity (A.P.I)	Desired Output/binary code
Paragneisses	2.65 - 2.85	5 - 15	70 - 130	100
Metabasites	2.75 - 3.1	5 - 20	0 - 50	010
Heterogeneous Series	2.60 - 2.9	1 - 15	40 – 90 & 120-190	001

Artificial Neural networks:

Artificial Neural networks (ANN) is a powerful tool for performing nonlinear functional mapping between a set of input variables (geophysical data) and a set of output or source parameters, together with particular procedures for optimizing the mapping (Bishop 1995).The importance of ANNs in this context is that they can function as universal approximators and are able to map nay continuous function to arbitrary accuracy(Yarger *et al.* 1978).This is achieved through adopting a massively parallel connectionist architecture of simple processing units (Perceptrons),the basic functioning of which was inspired originally by the biological neuron. The processing unit produces an output or is activated at a certain threshold determined by the value of its weighted input (Fig.1). The neural network optimizes the mapping by using a data set of training data, which contains examples of the functional mapping that the network is to learn (Raiche 1991). However, the main advantage of using neural networks to solve the inversion problem is that they effectively provide nonlinear mapping of the geophysical phenomena without assuming an explicit physical model of the process(Williams 1993).Therefore, the total time necessary for a neural network solution depends on the dimensions of the space of unknown parameters rather than the physical dimensions of the modeled area.(Spichak and Popova 2000).This makes ANNs very computationally efficient tools if multiple inversions are required, because once a network has been optimized or trained ,it effectively remembers the inversion solutions(Spichak and Popova 2000).It can therefore be applied easily to new or spatially extensive survey data with almost instantaneous results. However, for the effective processing of new data, it is important that the parameters of the training data used to train the original network be comparable with the new data. In this respect, the training data set must be designed carefully for maximum flexibility (Bescoby *et al.* 2006).

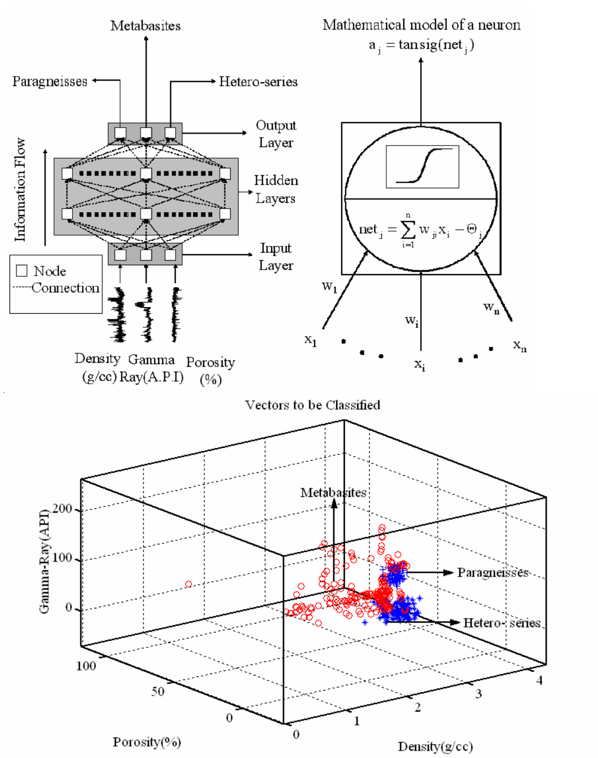

Figure 1: Topology of neural network and response function of a node (left), and the Perceptron representation of input and target vector clearly shows nonlinearity (right)(Maiti *et al*. 2007).

Artificial neural network (ANN) is also referred to Multi Layer Perceptron (MLP) network. Computational details of ANN could be made clearer by using some simple mathematics. In conventional approach of neural network learning a network is trained using a data set $S = \{x_j, D_j\}^N{}_{j=1}$ by adjusting network parameters w (weight and biases) so as to minimize an error function, such as (Aires 2004, Khan and Coulibaly 2006),

$$E_S = \frac{1}{2}\sum_{j=1}^{N}(D_j - O_j(x_j;w))^2 \quad \text{................(1)}$$

This objective function is a sum of terms, one for each input/target pair $\{x, D\}$, measuring how close the network output $O(x;w)$ is to the target D. The minimization of the error is based on iterative process of gradient E_S using "back propagation" (Rumelhart *et al.* 1986) .Often regularization is also included to modify the objective function:

$$E(w) = \mu E_S + \lambda E_R \quad \text{.....................(2)}$$

Where, for example, $E_R = \frac{1}{2}\sum_{i=1}^{R} w_i^2$, w is total number of weights and biases in the network. Experimental study suggests that the regularization term favors small values for network weight and biases and decreases and tendency of a model to "over fit" noise in the training data. Here λ and μ , which control other parameters (synaptic weight and biases), are known as hyper parameters. In this traditional approach, the training of a network starts with a initial set of weights and biases and ended up with the single best set of weights and biases given the objective function is optimized.

Bayesian Neural Network Paradigm:

In the Bayesian approach, a suitable prior distribution, say $P(w)$ of weights is considered before observing the data instead of single set of weights. Bayes theorem is used for writing an expression of the posterior probability distribution for the weights, say $P(w \mid S)$ as below (Aires 2004, Khan and Coulibaly 2006),

$$P(w \mid S) = \frac{P(S \mid w)P(w)}{P(S)} \quad \text{.....................(3)}$$

Where, $P(S \mid w)$ is the data set likelihood function, and the denominator, $P(S)$), is a normalization factor. Integrating over the weight space we can obtain

$$P(S) = \int P(S \mid w)P(w)dw \quad \text{......................(4)}$$

The above equation (4) ensures that the left hand side of (3) gives unity when integrated over all weight space. Once the posterior has been calculated, every type of inference is made by integrating over that distribution. Therefore, in implementing Bayesian method, expressions for the prior distributions $P(w)$, and likelihood function $P(S \mid w)$ is needed. The prior distribution $P(w)$ which is not related with data can be expressed in terms of weight decay regularizer, E_R of the conventional learning method. For example, if a Gaussian prior is considered, the distribution can be written as an exponential of the form (Bishop 1995)

$$P(w) = \frac{1}{Q_R(\lambda)}\exp(-\lambda E_R) \quad \text{.....................(5)}$$

Where, $Q_R(\lambda)$ is a normalization factor and can be given by

$$Q_R(\lambda) = \int \exp(-\lambda E_R)dw \quad \text{......................(6)}$$

The equation (6) ensures that $\int P(w)dw = 1$ .The hyperparameter λ can be fixed or could be optimized as part of the training process .Here Gaussian prior is preferred because it simplifies the calculation of the normalization coefficient $Q_R(\lambda)$ using equation (6) and gives

$$Q_R(\lambda) = \left(\frac{2\pi}{\lambda}\right)^{R/2} \quad \text{...............................(7)}$$

Alternative choices of prior $P(w)$ have been discussed by Buntine and Weigend (1991); Neal (1993) and Williams (1995).

The data dependent likelihood function in Bayes theorem can be written in terms of error function, E_S of the conventional method. For instance, if the noise (error) model is Gaussian, the equation for the likelihood function can be written as (Aires 2004, Khan and Coulibaly 2006),

$$P(S \mid w) = \frac{1}{Q_S(\mu)} \exp(-\mu E_S) \text{..........(8)}$$

The function $Q_S(\mu)$ is a normalization factor given by

$$Q_S(\mu) = \int \exp(-\mu E_S) dS \text{....................(9)}$$

Where, $\int dS = \int dD_1 .... dD_N$ represents an integration over the target variables. if it is assumed that the target data is generated from a smooth function with additive zero mean Gaussian noise, the probability of observing a data value D for a given input vector x would be.

$$P(D \mid x, w) \propto \exp\left(-\frac{\mu}{2}\{D - O(x;w)\}^2\right) \text{.........(10)}$$

Where $O(x;w)$ represents a network function governing the mean of the distributions, w represents the corresponding weight vector and the parameter μ controls the variance of the noise. Provided the data points are drawn independently from this distribution, we have the expression for the likelihood as (Aires 2004, Khan and Coulibaly. 2006),

$$P(S \mid w) = \prod_{j=1}^{N} P(D_j \mid x_j, w) = \frac{1}{Q_S(\mu)} \exp\left(-\frac{\mu}{2} \sum_{j=1}^{N} \{D_j - O_j(x_j;w)\}^2\right) \text{.....(11)}$$

The expression (9) for the normalization factor $Q_S(\mu)$ is then the product of N independent Gaussian integrals, which have been evaluated by Bishop (1995) and can be expressed as

$$Q_S(\mu) = \left(\frac{2\pi}{\mu}\right)^{N/2} \text{...(12)}$$

After deriving the expressions for the prior and likelihood functions, the posterior distribution of weights can be obtained using those expressions in eq (3) (Aires 2004, Khan and Coulibaly 2006),

$$P(w \mid S) = \frac{1}{Q_E} \exp(-\mu E_S - \lambda E_R) = \frac{1}{Q_E} \exp(-E(w)) \text{..............................(13)}$$

Where

$$E(w) = \mu E_S + \lambda E_R = \frac{\mu}{2} \sum_{j=1}^{N} \{D_j - O_j(x_j;w)\}^2 + \frac{\lambda}{2} \sum_{i=1}^{R} w_i^2 \text{......................(14)}$$

and

$$Q_E(\lambda, \mu) = \int \exp(-\mu E_S - \lambda E_R) dw \text{...........................(15)}$$

In equation (13), the objective function in the Bayesian method corresponds to the inference of the posterior distributions of the network parameters w .After defining the posterior distributions, the network is trained with a suitable optimization algorithm to minimize the error function $E(w)$ or to maximize the posterior distribution $P(w \mid S)$ .Using the rules of conditional probability, the distribution of outputs for a given vector, x ,can be written in the form (Aires 2004, Khan and Coulibaly 2006),

$$P(D \mid x, S) = \int P(D \mid x, w) P(w \mid S) dw \text{...........................(16)}$$

It may be noted that $P(D \mid x, w)$ is simply the model for the distribution of noise on the target data for a fixed value of the weight vector W_{MLP} and is given by (10), and $P(w \mid S)$ is the posterior distribution of weight. If the data set is large, the posterior distribution $P(w \mid S)$ may be approximated to a Gaussian distribution (Walker 1969).After some simplification, the integral in (16) can be written after Bishop (1995) as

$$P(D \mid x, S) = \frac{1}{(2\pi\delta_i^2)^{1/2}} \exp\left(-\frac{((D - O(x;w_{MLP}))^2}{2\delta_i^2}\right) \text{.....(17)}$$

and mean and variance is given by $O(x;w_{MLP})$, and

$$\delta_i^2 = \frac{1}{\mu} + g^T H^{-1} g \text{..................(18)}$$

Where μ is a hyper parameter, which is actually the inverse variance of the noise model, and g denotes the gradient of $O(x;w)$ with respect to the weights w evaluated at w_{MLP} and H is the Hessian matrix of the total(regularized) error function with elements given by (Aires 2004, Khan and Coulibaly 2006),

$$H = \nabla\nabla E(w_{MLP})$$
$$= \mu \nabla\nabla E_S(w_{MLP}) + \lambda I \text{...(19)}$$

Where, I is identity matrix. The standard deviation δ_i of the predicted distribution for the target D can be interpreted as error bar on the mean value $O(x;w_{MLP})$ .This error bar represents contributions from two sources, one is from the intrinsic noise on the target data, which is represented by the first term (18) and the other one is from the width of the posterior distribution of the network weights, corresponds to the second term in (18)(Aires 2004, Khan and Coulibaly 2006).The practical numerical experiment has been conducted following Nabney (2004).

Examples

Bayesian Neural Network are successfully applied to three sets of German Continental Deep Drilling Program (KTB) bore hole data e.g. density, neutron porosity and gamma ray log to classify the lithofacies succession and then was compared with our earlier results based on SSABP neural networks (see Maiti *et al.* 2007). The comparative results of both neural network based techniques for the pilot bore hole data are displayed in figures 2. Figure (2) exhibits the posterior probability distribution in a 3-coloumns gray-shaded matrix with black representing 1 and white representing 0. The maximum likelihood value corresponds to the class with maximum posterior probability: in ideal case, if the lithofacies of a particular class exists, the output value of the node in the last layer is 1 or very close to 1 and if not, it is 0 or very close to 0.

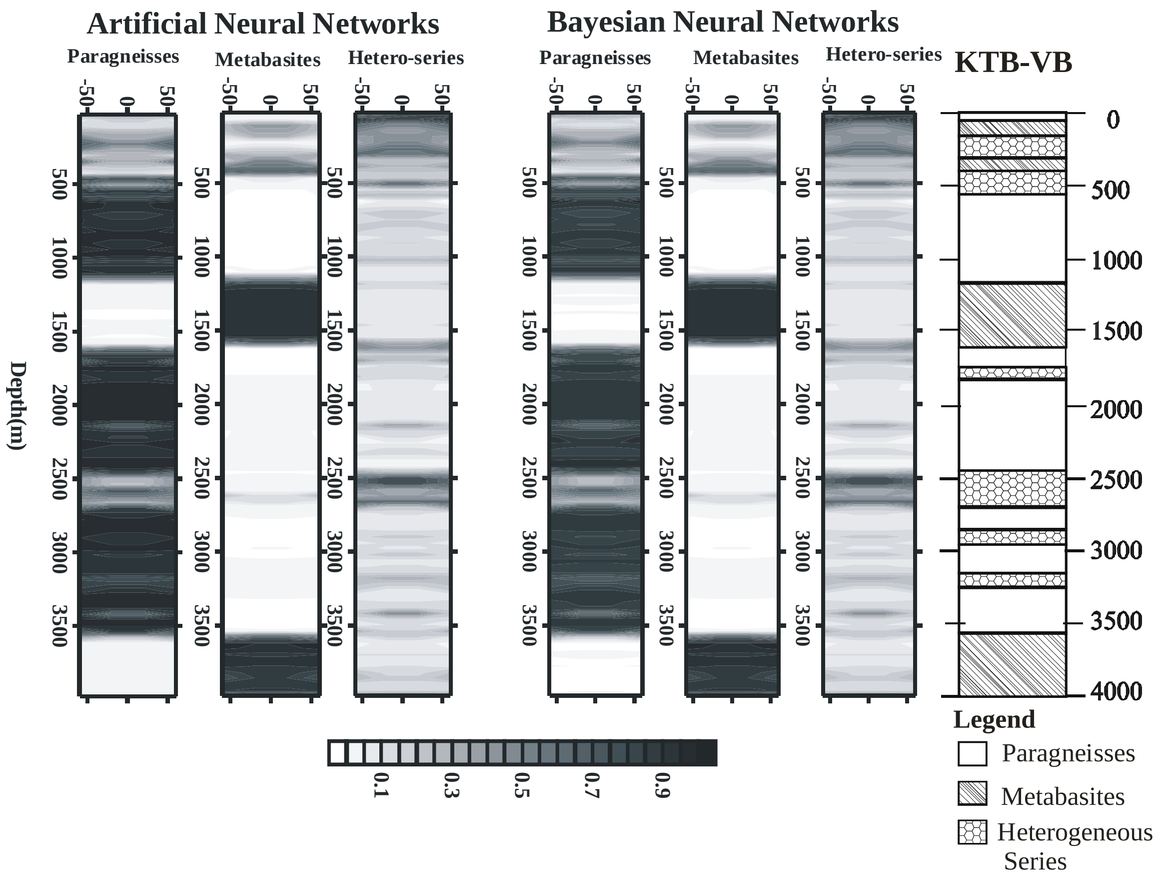

Figure 2: Comparision of ANN and BNN results with published lithofacies section of KTB pilot hole (KTB-VB)

Conclusions

Neural network program on Bayesian paradigm is developed and successfully applied to classify the litho-facies boundaries by using the well log data from the German Continental Deep Drilling Project (KTB). The Bayesian neural network (BNN) technique is an efficient and cost-effective tool to interpret large amount of borehole log data. The BNN based technique is also robust to analyze modestly correlated noisy data. A comparative study of the present model results with the published result suggests that the BNN method is also able to model the succession of some finer structures, which were hitherto not recognized. Such findings may have implications in understanding the crustal inhomogeneity. Because of its computational efficiency of Bayesian learning, in terms of uncertainty analysis, capability to inherently taken care of overfitting and underfitting problem, it is proposed that the present methods can be further exploited for analyzing large number of borehole data in other areas of interest. Finally, some deviations observed in BNN neural network analysis from the prior knowledge seem to be interesting and should provide a basis for more detailed examination of the geological significance of finer structures intervening bigger geological units.

Acknowledgements

Saumen Maiti expresses sincere thanks to Department of Science and Technology (DST), Govt. of India to carry out the research work**.** SM is also thankful to Prof. Archana Bhattacharyya, Director, Indian Institute of Geomagnetism (IIG), Navi-Mumbai for her kind permission to publish this work. SM expresses sincere thanks to Prof. S. G. Gokaran, and Dr.G.Gupta, IIG, Navi-Mumbai for helpful discussions. We are also thankful to Prof. Hans-Joachim Kumpel for providing the data.

References

Aires, F., 2004.Neural network uncertainty assessment using Bayesian statistics with application to remote sening: 1.Network weights,J.Geophys.Res.,109,D10303,doi:10. 1029/2003JD004173

Berckhemer, H., Rauen, A, Winter, H.,Kern, H., Kontny, A., Lienert, M., Nover, G., Pohl, J., Popp, T., Schult,A.,Zinke, J., and Soffel, H.C.,1997.Petrophysical properties of the 9-km deep crustal selection at KTB.J.Geophys.Res.,102,no.B8,pp.18337-18361.

Bescoby, D.J., Cawley, G.C.,and Chroston, P.N.,2006.Enhanced interpretation of magnetic survey data from archaelogical sites using artificial neural networks,Geophysics,71,5,H45-H53.

Bishop C.M., 1995.Neural Networks for Pattern Recognition. Oxford University Press

Buntine, W.L., and Weigend, A.S., 1991, Bayesian back propagation, Complex Syst, 5, 6030643

Busch, J.M., Fortney, W.G., and Berry, L.N. 1987.Determination of lithology from well logs by statistical analysis: SPE Formation Evaluation, 2,412-418

Coulibaly, P., Anctil, F. and Bobee, B 2001a, Multivariate reservoir inflow forecasting using temporal neural networks, J,Hydrol.Eng.6,367-376

Delfiner, P.,Peyret,O.,and Serra, O.,1987.Automatic determination of lithology from well logs:SPE Formation Evaluation,2,303-310

Emmermann, R., and Lauterjung, J., 1997.The German Continental Deep Drilling Program KTB:Overview and major results. J. Geophys. Res., 102, .18179-18201

Franke, W., 1989.The geological framework of the KTB drill site. In: Emmermann, R., Wohlenberg,J.,(Eds.), The German Continental Deep Drilling Program (KTB).Springer, Berlin, pp.38-54

Gassaway, G.R.,DRMiller,LE Bennett,RA Brown M.Rapp and V.Nelson,1989,Amplitude variations with Offset:Fundamentals and Case Histories,SEG Continuing Education Course Notes

Khan M.S., and Coulibaly P., 2006. Bayesian neural network for rainfall-runoff modeling, Water Resour. Res. 42, W07409,doi:10.1029/2005WR003971

Leonardi, S., and Kumpel, H., 1998.Variability of geophysical log data and signature of crustal heterogeneities at the KTB. Geophys. J. Int., 135,964-974.

Leonardi, S., and Kumpel, H., 1999. Fractal variability in super deep borehole-implications for the signature of crustal heterogeneities. Tectonophysics, 301, 173-181

Maiti, S., and Tiwari, R.K., 2005: An Automatic Method for Detecting Litholog Boundary using Walsh Transform: A Case Study from KTB Borehole. Computers and Geosciences, vol 31, 8, pp.949-955

Maiti, S., and Tiwari, R.K.,.2005 Identifying Litho-facies Boundaries using Super Self Adaptive Back Propagation Neural network (SSAB): A Case Study from the KTB Borehole, Petrotech, Delhi, Extended Abstract

Maiti, S.,Tiwari,R.K., and Kumpel H.J ,2007.Neural network modeling and classification of lithofacies using well log data: a case study from KTB borehole site,Geophys.J.Int.169,733-746

Nabney, I.T2004, Netlab Algorithms for pattern recognition ,Springer,New York

Neal,R.M,1993,Bayesian learning via stochastic dynamics in Advances in neural information Processing Systems 5, edited by C.L.Giles et al.,475-482,Elsevier,New York

Peching, P., Haverkamp, S., Wohlenberg, J., Zimmermann, G., and Burkhardt, H., 1997.Integrated interpretation in the German Continental Deep Drilling Program: Lithology, porosity, and fracture zones. J. Geophys. Res.,102, 18363-18390

Pickett, G.R., 1963,Acoustic character logs and their application in formation evaluation.J.Petr.Tech, 15,659-667

Raiche, A., 1991.A pattern recognition approach to geophysical inversion using neural nets. Geophys.J. Int., 105,629-648

Rogers, S.J., Fang, J.H., Karr, C.L., and Stanley D.A., 1992. Determination of lithology from well logs using a neural network. Am Ass. Petrol. Geol. Bull., 76, no.5, 731-739

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. Learning representations by back-propagating errors. Nature, 323, 533-536.

Spichak,V., and Popova,I.,2000.Artificial neural network inversion of Magnetotelluric data in terms of three-dimensional earth macroparameters,Geophys.J.Int 142,15-26

Walker, A.M,1969,On the asymptotic behaviour of posterior distributions,J.R.Stat.,31,80-88.

Wolff, M., and Pelissier-Combescure,J.,1982,FACIOLOG:automatic electrofacies determination:SPWLA Annual Logging Symposium paper FF,6-9

Williams, P.M., 1995,Bayesian regularization and pruning using a Laplace prior,Neural Comput.,7,117-143

Yarger,H.L.,Robertson,R.R.,and Wentland, R.L.1978,Diurnal drift removal from aeromagnetic data using least squares,Geophysics,46,1148-1156