8th Biennial International Conference
& Exposition on Petroleum Geophysics

"HYDERABAD 2010"

P-188

Automatic detection of Litho-Facies via the Hybrid Monte Carlo Based Bayesian Neural Networks Approach

[1]Saumen Maiti[*], and [2]Ram Krishna Tiwari
[1]*Indian Institute of Geomagnetism, Navi-Mumbai, India.*
[2]*National Geophysical Research Institute, Hyderabad, India.*

Summary

The precise litho-facies change from well log records is a complex non-linear problem in geophysical data processing. Recorded well log signals are a complex superposition of non-stationary/non-linear signals of varying wavelengths and frequencies, shaped by the heterogeneous composition and structural variation of rock types in the earth. This weakens our ability to use traditional statistical techniques, which either fail in most cases to discriminate and/or, at best, do not precisely extract facies changes from such a complex well log signals. We propose here a new method, set in a Bayesian neural network (BNN) framework and employing a powerful Hybrid Monte Carlo (HMC)/Markov Chain Monte Carlo (MCMC) simulation scheme to identify facies changes, from complex well log data. We first construct a complex, composite, synthetic time series using the data from three simple models,i.e., first order auto-regressive Ar(1), logistic, and random white noise, and then we attempt to identify individual signals in the pooled synthetic time series. We use the autocorrelation and the spectral characteristics of the individual signals as input vectors for training, validating, and testing the Artificial Neural Network (ANN) model. The results show that the Bayesian separation scheme provides consistently good results, with accuracy at more than 74%. The method is demonstrated using well log data from the German Continental Deep Drilling Program (KTB). The scheme is able to discriminate lithofacies boundaries with accuracy of (~92%) in validation and of (~93%) in test samples. The efficacy of the BNN in the presence of colored noise suggest that the designed network topology is robust for up to 30% correlated noise; however, adding more noise (say 50% or more) obscures the desired signals. Our method provides a robust means for decoding finely detailed successions of lithofacies from complex well log data, yielding and understanding of the nature of the underlying inhomogeneous crust.

Introduction

It is accepted that sharp changes in rock properties recorded in form of well logs reflect physical boundaries and facies changes within various rock types. The well logs recorded in such rocks exhibit complex signal characteristic comprising nonlinear/non-stationary and random behavior. Discrimination of different rocks types/facies changes from such complex well log signals is, therefore, one of the important problems in geophysical signal analysis. During the past several years, conventional method such as graphical cross-plotting, and other statistical techniques e.g. multivariate statistical methods such as principle component and cluster analyses (Wolff & Pelissier-Combescure 1982), and discriminant function analysis (Delfiner et al. 1987) have frequently been used to study borehole data. However, in complex geologic situations, such as in the presence of crystalline rocks where metamorphism leads to facies changes, it is not easy to extract accurate information from well log data using these conventional methods. Moreover, inferences drawn by such methods are also found to be ambiguous because of strong overlapping of nonlinear/non-stationary well log signals, which are also found to be tainted with deceptive, colored noise. The traditional techniques are semi-automated and require a large amount of data, which are costly and not always easily available. Further, these methods are also very tedious and time-consuming, particularly when dealing with noisy, complex borehole data.

The Artificial Neural Network (ANN) technique is used extensively for the classification of complex nonlinear signals, due to its inherent ability to approximate the functional relationship between the input and the output space/domain simply by learning through examples, even if

E-mail: saumen_maiti2002@yahoo.co.in.

there is no deterministic relationship between the input and the output space/domain (Bishop, 1995). ANN has consequently been applied to almost all branches of geophysics. Recently, Maiti and Tiwari have developed multiple linear and nonlinear algorithms for identification of rock boundary from the KTB borehole signal (Maiti and Tiwari 2005; Maiti et al., 2007). However, one of the problems with the very popular ANN-based back-propagation algorithm (Rumelhurt, 1986) is that it does not converge to global minimum during the optimization process. We, therefore, propose a more powerful approach known as "Bayesian Inference" (Tarantola, 1987) to approximate the posterior probability distribution from data likelihood and prior information using a Monte Carlo algorithm in the case of the KTB well log data. These methods have proven to be very useful in several other contexts because they yield non-unique solutions of complex geophysical inverse problems. The practical use of a sampling-based inversion scheme, e.g., Hybrid Monte Carlo (HMC)/Markov Chain Monte Carlo (MCMC), for neural network training can be found in several previous works (Neal, 1996; Bishop, 1995). More recently, Maiti and Tiwari (2009) has developed new algorithm for borehole data analysis. However, the efficacy and applicability of these theoretical developments are not well explored for the case of complex and noisy signals. Hence, we explored the stability of the method on noisy, synthetic, nonlinear models of varying complexity and then applied the method to the real KTB data. We tested the method on various synthetic data generated from the well known models: (1) Ar (1)/ stochastic random walk model, (2) complex/logistic, and (3) white noise. This experiment was intended to provide some useful guidelines and confidence in classifying complex data sets and thereby help in making sound physical interpretations of actual well log data. We applied the method to the following KTB bore hole data: (i) density (RHOB), (ii) neutron porosity (NPHI), (iii) gamma ray intensity (SGR), (iv) seismic P-wave transit travel time (DTCO), and (v) electrical resistivity (LLD) to discriminate among three litho-facies in a complex metamorphic region of central Europe.

Data and Method

Multilayer Perceptrons

Multilayer Perceptrons (MLPs) networks are parallel computational units composed of many simple processing elements which mimic the biological neurons (Maiti and Tiwari 2009). Processing elements/ nodes are interconnected layer by layer and the functions of the each node are determined by connections, weights and biases, and the topology of the network (Bishop, 1995; Maiti et al 2007; Maiti and Tiwari 2009). In the popular back-propagation method, the error is usually minimized by adjusting the weights and biases using a gradient-based iterative chain rule from output to input layer (Rumelhurt, 1986). The main drawback of this method is that it often becomes stuck in local minima on an error surface.

In order to avoid the latter problem, we use hybrid Monte Carlo simulations (also well known as a leapfrog discretization scheme) in conjunction with Bayesian probability theory, which is naturally parsimonious, thus suiting our present needs. The complete details of the neural network topology and the learning rules can be found elsewhere (Poultron, 2001; Bishop, 1995). Here we can relate the geophysical observations to the model through the following forward equation:

$$x = f(d) + \varepsilon \quad \text{..... (1)}$$

where f is a nonlinear function relating the model space and data space, ε is error, and x d are the data and the model respectively. A common way of inverting for the model, d in equation 1 is via iterative least squares method. This, however, does not provide uncertainty measures, which are essential for sound physical interpretation of geophysical observations (Tarantola, 1987).

Model Solution in the Bayesian frame work

To solve the above equation 1 in the Bayesian framework, we recast it as follow

$$d = f_{NN}(x; w) \quad \text{..... (2)}$$

where, f_{NN} is output predicted by the network and w are the network weight parameters. In the conventional approach for solving equation 1, regularization is often included to minimize the misfit function:

$$E(w) = \mu E_S + \lambda E_R \quad \text{..... (3)}$$

where, $E_S = \frac{1}{2}\sum_{k}^{N}\{d_k - o_k(x_k; w_k)\}^2$ and, $E_R = \frac{1}{2}\sum_{i=1}^{R} w_i^2$.

R is the total number of weights and biases in the network, while λ and μ , which control other parameters (synaptic weight and biases), are known as hyper parameters. In this approach, the training of a network starts with an initial set of weights and biases and ends up with the single best set of weights and biases to optimize the objective function. In the Bayesian approach, a suitable prior distribution, say $P(w)$ of weights is considered before observing the data, instead of considering only a single set of weights. Using Bayes' rule, an *a posteriori* probability distribution for the weights, say $P(w \mid s)$ can be defined as (Khan and Coulibaly, 2006),

$$P(w \mid s) = \frac{P(s \mid w)P(w)}{P(s)} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

Here, $P(s \mid w)$ is a data set likelihood function, and the denominator, $P(s)$ is a normalization factor. The denominator $P(s)$ is intractable, so direct estimation of posterior $P(w \mid s)$ is not possible. Using the rule of conditional probability, the distribution of the output for a given input vector x can be defined in the form (Khan and Coulibaly, 2006),

$$P(d \mid x, s) = \int P(d \mid x, w)P(w \mid s)dw \quad \ldots\ldots\ldots\ldots (5)$$

The major problem in Bayesian computation is evaluating the integrals for the posterior weights (equation 4) and for the network output (equation 5). In this regard, the MCMC sampling-based method plays an important role in evaluating posterior integrals. Equation 5 can be approximated as

$$P(d \mid x, s) = \frac{1}{N}\sum_{n=1}^{N} P(d \mid x, w_m) \quad \ldots\ldots\ldots\ldots\ldots (6).$$

where $\{w_m\}$ represents a MCMC sample of weight vectors obtained from the distributions $P(w \mid s)$, and N is the number of points w sampled from $P(w \mid s)$.

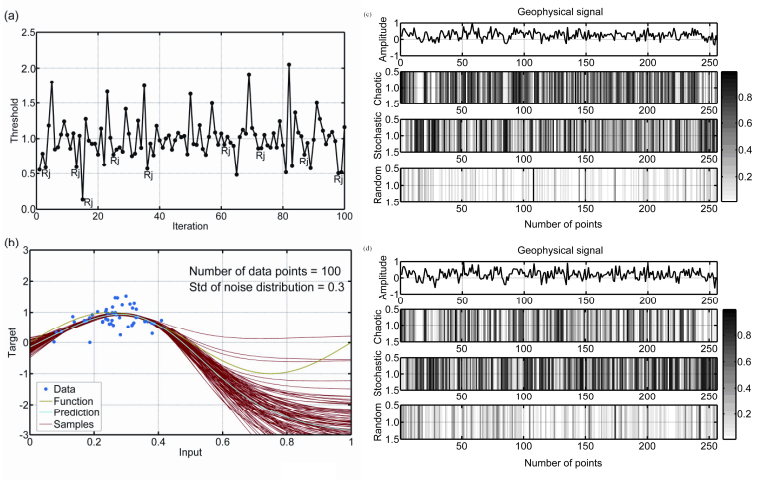

Figure 1: (a) The acceptance threshold history against iteration. Rj denotes the rejected threshold. (b)The graph shows the simulation of a synthetic underlying function using 100 numbers of data points in the presence of noise via Hybrid Monte Carlo method. The standard deviation (std) of the noise is 0.3.(c) HMC-based classification results of composite signals. Here in the chaotic model the value for constant "B" is 3.8, and for the stochastic model the assigned value for "A" is 0.5. Random white noise signal are added point by point (d). Same as above with "B" value = 4.0 for chaotic model and "A"=0.25 for stochastic model and random white noise signal as above.

Hybrid Monte Carlo (HMC)

In the HMC algorithm, each trajectory is updated by approximating the Hamiltonian differential equations by a leapfrog discretization scheme (Duane et al., 1987). The MCMC algorithm draws an independent and identically distributed (i.i.d) sample $\{w^{(i)}; i = 1,2,....N\}$ from the target distribution $P(w|s)$. The Markov process forms a sequence of "states" to draw samples from the posterior probability. The chain converges to $P(w|s)$ if given enough space to do so. The complete mathematical details can be found elsewhere (see Bishop, 1995; Nabney, 2004). It may be noted, however, that the pure Metropolis-Hastings algorithm (Hastings, 1970) is found to be very slow because the method makes no use of gradient information. Contrary to this, the HMC based algorithm for sampling from the target distribution makes use of gradient information. The following steps are needed once a step size θ and the number of iterations L have been decided upon: (i) randomly choose a direction τ: τ can be either -1 or +1 with probability 0.5 to simulate a forward or backward step in time; (ii) following the theory of Hamiltonian statistical mechanics, the transition probability matrix should satisfy microscopic reversibility, which means that probability of these two transitions from q_j to q_i or from q_i to q_j be same at all times and each pairs of points maintains a mutual equilibrium (iii) iterate, starting with the current state $[q, p] = [(q(0), p(0)]$ of energy H, where p is a momentum term which is randomly evaluated at each step; and (iv) let algorithm be applied L times with a step size of θ resulting in the candidate state, $[w^*, p^*]$ with energy H^*.The candidate state is accepted with usual Metropolis probability of acceptance, $\min\{1, \exp[-(H^* - H)]\}$ (Bishop, 1995). If the candidate state is rejected then the new state will be the old state. These steps, in essence, describe how the sampling is done from a posterior distribution of network parameters so that the summation of equation 6 can be accomplished and the posterior distribution can be found, thus allowing the optimization of the network. The main idea of the algorithm is that the acceptance probability is evaluated at each step so as to produce the necessary number of realizations . Thus we estimate the posterior distributions which are not easily tractable analytically. The desired statistics can easily be estimated from the same realizations.

Examples

Construction of the synthetic data

Before these methods can be used on actual well log data, we test the efficiency of the proposed techniques on the following three theoretical models: (i) Ar(1)/ stochastic/ random walk model, (ii) logistic, and (iii) a white noise process. A brief description of these models is given below:

Stochastic model / auto-regressive (1) model

A first order autoregressive Ar (1) model/random walk model (Maiti and Tiwari 2009) takes the form $X_t = AX_{t-1} + \varepsilon_t$; where $t = 1,2,3,.....N$ denotes the discrete spatial increment. Here A is a maximum-likelihood estimator and another parameter ε_t denotes a purely random process (an uncorrelated normal distribution uniformly distributed in the interval between 0 and 1). An autocorrelation coefficient describes the degree of signal correlation in the noise and is calculated from the data. It has a value ranging from 0 to 1. X_t depends partly on X_{t-1} and partly on the random distribution ε_t. The Ar (1) / stochastic model exhibit a tendency to cluster towards low values (Maiti and Tiwari 2009).

Logistic model

A complex system can be represented by the "logistic" model equation which is of the form of $X_{t+1} = BX_t(1 - X_t)$; Where X_t and X_{t+1} are the present and future values of generating process with relative values ranging from 0 to 1, and B is a coefficient (control parameter) between 0 and 4. Theoretically, complexity reaches the maximum for a B value ~4.

Random white noise model

Random white noise is uncorrelated and has zero mean. Such a process is unpredictable due to its uncorrelated nature. A random 3-D phase plot shows that the values scatter equally in all directions (Maiti and Tiwari 2009).

Network model for synthetic data

We train the MLP network coupled with HMC using the data of the three basic models as described above, their power spectra and ACF coefficients. In all, 768 data sequences were used for network training. A composite time series of three models (Figure 1a-b) is used to test the trained network. Spectral discriminates are used to train the network, and the resulting network output for presence or absence of a nonlinear sequence is indicated in the form of probabilistic index (Maiti and Tiwari 2009). One can see that the trained network is able to discern individual signal from the test model sequence with 74% accuracy (Figures 1c-d). Uncertainty analysis of the predicted output is performed by calculating standard deviation (STD) from the posterior covariance matrix of the network output. The discrimination results are presented in 3 gray bands, with black representing 1; and white, 0 . We have experimented with different network parameters to estimate the uncertainty of the network output (Maiti and Tiwari 2009). The average uncertainty of the predictions is ~0.14 at network output, with a 90% confidence interval. We see that the uncertainty of the network prediction depends on the hyperparameter μ (see Maiti and Tiwari 2009). It is interesting that classification results for the composite signals (comprising non-linear, stochastic and random sequences) by the present method are very good. Thus, the proposed method could be used as an alternative way to discriminate among complex signals.

Table 1: Significant limits to generate forward model for neural network learning

Rock type	Density RHOB[g/cc]	Neutron porosity NPHI[%]	Gamma ray intensity SGR[A.P.I]	P-wave transit travel time DTCO($\mu s/m$)	Ln(Resistivity) LLD[ohm-m]	Desired output [binary code]
PG	2.65 - 2.85	5 - 15	70 - 130	165-200	3-9	100
MB	2.75 - 3.10	5 - 20	0 - 50	143-196	4-10	010
HS	2.60 - 2.90	1 - 15	40 – 90&120-190	174-220	5-9	001

Network model for field data

After describing the successful test of the proposed algorithm on complex synthetic series, we will demonstrate the method on the following real KTB, main-bore-hole records: density (RHOB), neutron porosity (NPHI), gamma ray intensity (SGR), seismic P-wave transit travel time (DTCO) and electrical resistivity (LLD) for discriminating among three litho-facies in a complex metamorphic region of central Europe.

Bayesian Network model for the KTB borehole data

We parameterized five sets of log response (considering the well log-response of RHOB, NPHI, SGR, DTCO and LLD to the rock compositions, and the significant parameter limit as described above and in Table 1 and generated corresponding representative input/output pairs. A total of 1408 input/target pairs were considered for the present analysis. The total available data sets were randomly partitioned for training (50%), for validations (25%) and for testing (25%).

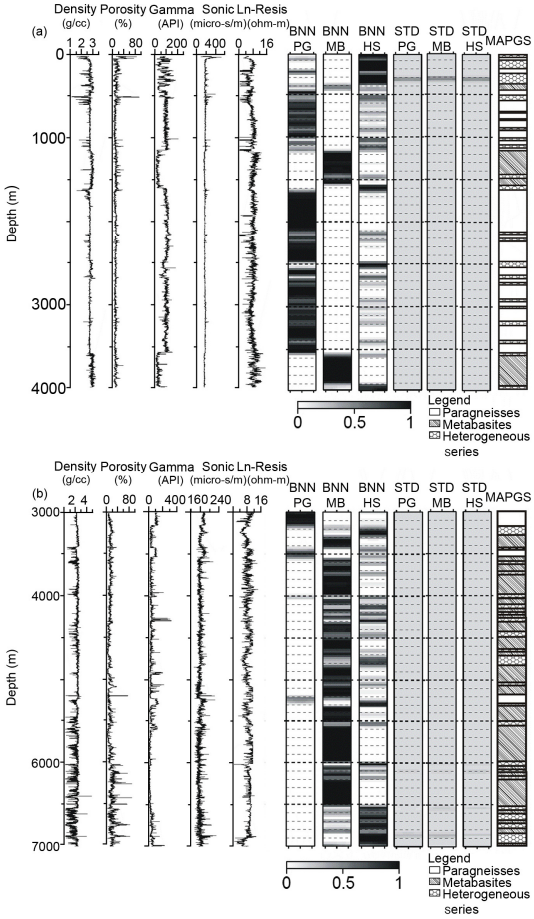

Figure 2: (a) Maximum *a posteriori* geological section (MAPGS) based on Bayesian neural networks with hybrid Monte Carlo simulations of pilot borehole (KTB-VB) from depth 28-4000m. (b) Same for main borehole (KTB-HB) from depth 3000-7000m.

We mention here, however, that there is no rule of thumb for such a partitioning of the data. Hence the choice of percentage of data divided among training, validation and testing is not unique, but rather it is data and problem dependent. We further re-checked different target units in each independent subset carefully and reshuffled the total samples (1408 input and output pairs) and partitioned them randomly to each independent subset. After successful completion of the training and the test for robustness of the proposed method, the trained network was used for discrimination among lithofacies in the pilot bore hole (KTB-VB) for a depth interval of 28-4000m and in the main hole (KTB-HB) for a depth interval of 3000-7000m. The pilot borehole and main hole data were sampled at a common interval of 0.1524m (6 inches). The total depth of the main hole and the pilot hole are 9101m and 4000m respectively. It is important to mention here that the well logs used here are continuous and uninterrupted throughout the depth intervals used.In Figure 2(a-b), the maximum a *posteriori* geological section (MAPGS) corresponds to the class with maximum a *posteriori* probability: in the 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. The (MAPGS) is consistent with main borehole lithosections (Figure 2, Maiti et al., 2007). In addition to this, the proposed method also detects some finer structures over a couple depth intervals within the major successions of paragneiss (PG), Metabasites(MB), and heterogeneous series(HS). These finer structures appear to be geologically significant (Maiti et al., 2007). These additional findings of HS in between the PG and MB were ambiguous in previous studies because of the strong superposition of well log signals produced by the varying composition and structure of the crystalline rocks. For example, at depth intervals of 500-600m, 1010-1040m,1100-1120m,1510-1550m,2500-2520m and 3400-3500m for the pilot hole (KTB-VB) and 3210-3260m, 4000-4010m, 4100-4150m, 4800-4810m, 5300-5310m, 5500-5520m, 6000-6050m, 6530-6550m, 6750-6800m, 6900-6960m for the main hole (KTB-HB), the HS litho-type is detected (Figure 2a-b). In the main hole, the change of MB to HS at depth 3413.15m is also confirmed. It is geologically significant in view of the extension of the Franconian Lineament which cuts the KTB main hole (KTB-HB) at that particular depth. The present analysis shows clear demarcation of the three types of lithofacies.

Conclusions

We have developed a new HMC-based Bayesian neural network method to discriminate among geophysical signals emanating from the complex geological sources. The important advantage of this HMC-based Bayesian neural network approach is that it can discriminate accurately complex signals even in the presence of different kinds of noises that are encountered in many geological situations.

Our KTB data analysis using the proposed HMC-based Bayesian neural network method suggests that the method is able to accurately classify litho species boundaries. Further, in addition to agreeing well with earlier findings our model result also suggests presence of some finer bed boundaries that were ambiguously missing in the previous studies. Presence of finer structures seems to have geological significance for understanding the crustal inhomogeneity and structural discontinuities within the central European crust. Thus our main contribution has been to demonstrate how the new HMC-based Bayesian neural network method could be used to generate lithofacies boundaries from the well log signals decoded from the borehole. The method could be further exploited for modeling different kinds of geological and geophysical well log signals.

Acknowledgements

We are grateful to Dr V.P. Dimri, Director, NGRI and Prof. Archana Bhattacharyya, Director, Indian Institute of Geomagnetism (IIG), Navi-Mumbai for their support and kind permission to publish this work. Saumen Maiti (SM) expresses sincere thanks to the Department of Science and Technology (DST), Govt. of India for support of the research work. We are also thankful to Prof. Hans-Joachim Kumpel for providing the KTB borehole data.

References

Bishop, C. M., 1995, Neural networks for pattern recognition: Oxford University Press.

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

Duane, S., A.D. Kennedy, B.Pendleton, and D. Roweth, 1987, Hybrid Monte Carlo: Physics Letters B, **195**, 216-222.

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

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov Chain and their applications: Biometrika, **57**, 97-109.

Leonardi, S., and H. Kumpel, 1998, Variability of geophysical log data and signature of crustal heterogeneities at the KTB: Geophysical Journal International, **135**, 964-974.

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

Maiti, S., and R.K.Tiwari, 2005, An automatic method for detecting litholog boundary using Walsh Transform: A case study from KTB borehole: Computers and Geosciences, **31**, 8, 949-955.

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

Maiti, S., and Tiwari, R.K., 2009, Automatic discriminations among geophysical signals via the Bayesian neural networks approach, Geophysics (In press)

Maiti, S., and Tiwari, R.K., 2009, A Hybrid Monte Carlo Method Based Artificial Neural Networks Approach for Rock Boundaries Identification: A Case Study from the KTB Bore Hole, Pure and Applied Geophysics, DOI 10.1007/s00024-009-0533-y

Nabney, I. T., 2004, Netlab Algorithms for pattern recognition: Springer, New York.

Neal, R.M., 1996, Bayesian learning for neural networks: Springer-Verlag, NewYork, Inc.

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

Tarantola, A., 1987, Inverse Problem Theory: Elsevier, New York.

Wolff, M., and J. Pelissier-Combescure, 1982, FACIOLOG:automatic electrofacies determination: Society of Petrophysicists and Well log Analysts Annual Logging Symposium paper FF,6-9.