Application of Attribute Analysis and Probabilistic Neural Network to predict Missing Log data

Abir Banerjee[1], Rima Chatterjee[2], A.K.Seth[1]*
[1]ONGC Ltd. Bokaro, India, [2]IIT-ISM Dhanbad, India

Email: *Banerjee_Abir.ongc.co.in*

Keywords

Attribute analysis, Regression equations, Probabilistic Neural Network (PNN).

Abstract

Petro-Physical parameters obtained from the well logging data is crucial in decision making for reserve estimation. Tool response is an important criteria in the acquisition of good quality data. Malfunctioning of logging tool leads to data loss while recording the data. To overcome this problem, a methodology has been developed to generate the missing log from the other nearby wells log data. In one well of CBM block located in Bokaro field the density log data could not be recorded due to tool malfunctioning. The developed methodology is followed to predict the missing data using seven nearby well data. The density data is selected as the target log with the training of external attributes. Single attributes analysis is carried out and number of regression equations are generated. The attribute with minimum error and maximum correlation relationship is selected as input to initiate the multi-attribute analysis. Multi-attribute analysis are carried out by creating a list of transform by stepwise regression between a subset of the attributes and the target log values from single attribute. Proper gain and lag are applied by Probabilistic Neural Networks (PNNs) using the concept of distance in attribute space from known to unknown point. Attribute with minimum value in training and validation error is selected to obtain the relationship between the predicted and observed density data to generate the final data. The predicted density data is validated with nearby well data. The result obtained using multi attribute and PNN has correlation 149.2 and error 0.91 compared to single attribute correlation 0.80 and error 214.5. This methodology of artificial intelligence helps to predict the missing log with accuracy and to obtain sub-surface information.

1. Introduction

The sub-surface information can be accurate analyzed by acquiring logging data using different sensors inside the well. The study area is located in CBM block of Bokaro field in the state of Jharkhand, India. The eight nearby wells numbered W-1, W-2, W-3, W-4, W-5, W-6, W-7 and W-8 as shown in the study area Figure-1, where well W-6 has the density log missing due to tool malfunctioning during data acquisition. In order to regenerate the density data the prediction of missing log is necessary to obtain the petro-physical information in the well where crucial information is hiding. Attribute analysis approach is widely used in oil and gas industry for the identification of large numbers of parameters as a problem or a collection of many problem in pattern recognition. The basis for the approach in attribute analysis is to identify the large number of objectives which can be combined in variety of ways or pattern analysis and generation of regression equations with best regression co-efficient (*Justice et al.; 1985*). PNN is the pattern classification algorithm based on the concept of the identification of nearest neighbor and gives weight according to the neighbor distance in multi-dimensional attribute space. The weighted PNN is more robust approach nowadays used which measures the performance of the training set, restrict the use of parameters which increases the statistical significance and optimize the parameter of the metric using non-linear search techniques (*Atkeson 1991; Kelly and Davis 1991*).

The objective of the paper is to highlight and discuss the developed methodology for prediction of the missing density log data in one well from the data available of seven nearby wells using the application of attribute analysis with linear and non-linear transform by stepwise regression and PNN and to compare the error and correlations results of single with multi-attribute analysis.

Application of Attribute Analysis and Probabilistic Neural Network to predict Missing Log data

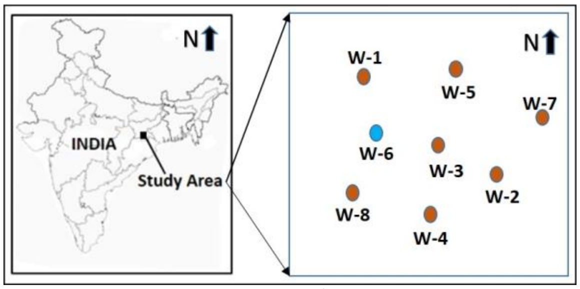

Figure 1: Representational Image of the study area with the well location (not in scale), well W-6 has missing density log data.

2. Methodology and Attributes

The methodology workflow to predict the missing log is shown in the Figure-2. Missing density log data is selected as the target attribute that is needed to be predicted. Attributes are classified as external and internal, where external attributes are the data present in wells like Gamma Ray (GR), Latero Log Deep (LLD), Latero Log Shallow (LLS) and Neutron porosity (NPHI) and internal attributes are the modified external attributes by applying linear or non-linear transform. Target well W-6, and nearby seven wells are selected to train the data collectively with specified sampling rate of 0.125m to establish the data list for the prediction of target density log (*Hampson Russell et. al., 2001*). The training plot is shown in Figure-3, the values of target density log, external attributes are sampled at different depth interval in the analysis window zone and categorized.

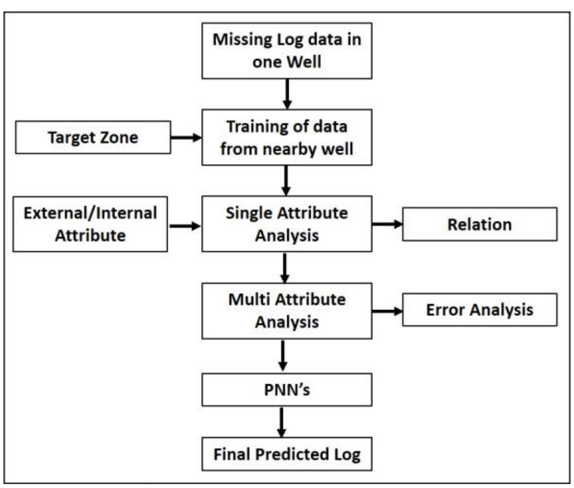

Figure 2: Workflow to predict the missing log.

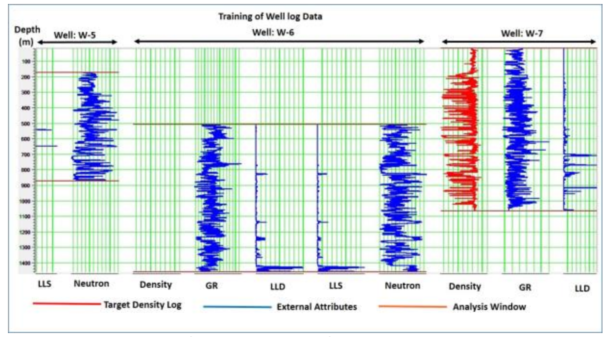

Figure 3: Training of well log data of all wells (W-1 to W-8).

3. Single Attribute Analysis

Single Attribute analysis is carried out to facilitate the selection and inspection of the relationship between target density log and external or internal attributes. Number of attributes are generated between target log and external attributes. Various transform are applied to target log and external attributes to make non-linear to linear transform and obtaining new internal generated attributes from external attributes. Table-1 shows the list of few single attribute analysis with error and correlation results out of 100 samples of generated attributes. Out of 100 samples of generated attributes, the one with minimum error (214.51) and maximum correlation (-0.80) (Attribute-4: Density^2 and Neutron porosity) is selected for the prediction of density log by single attribute regression shown in Figure-4. The predicted density log of well W-6 is represented in Figure-5 by applying the regression equation by single attribute analysis.

S. No.	Target	Attribute	Error	Correlation
1	Density^2	NPHI	214.51	-0.80
2	Density	NPHI	222.49	-0.80
3	Density	Sqrt [NPHI]	250.32	-0.74
4	Log [Density]	Log [LLD]	312.18	-0.66
5	Density	LLD	350.82	-0.34

Table 1: List of few Single attribute analysis with error and

correlation result out of 100 samples of generated attributes.

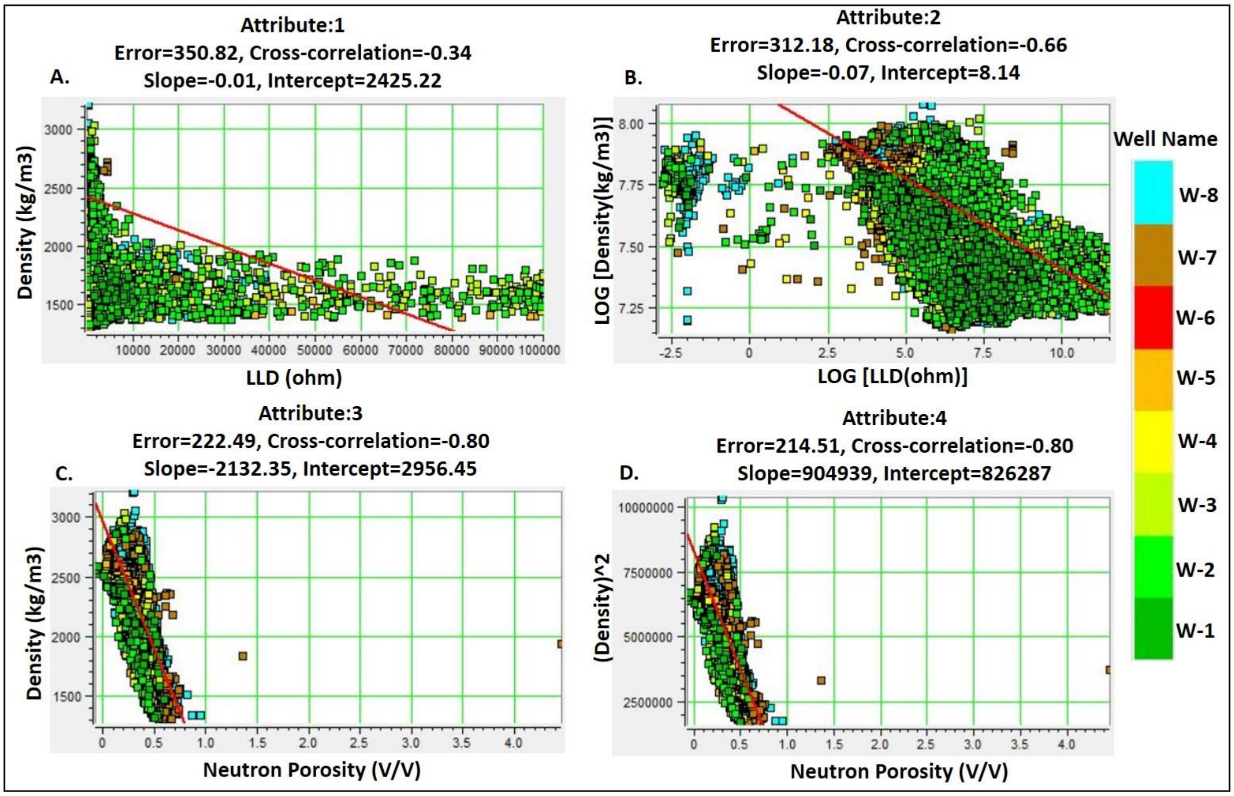

Figure 4: Single attribute analysis result using different attribute. A. Attribute: 1 is the LLD with non-linear transform. B. Attribute: 2 is the Logarithmic transform of LLD with linear transform. C. Attribute: 3 is the Neutron Porosity. Attribute 4: is the Neutron Porosity with square of the density with minimum error and maximum correlation out of 100 attributes.

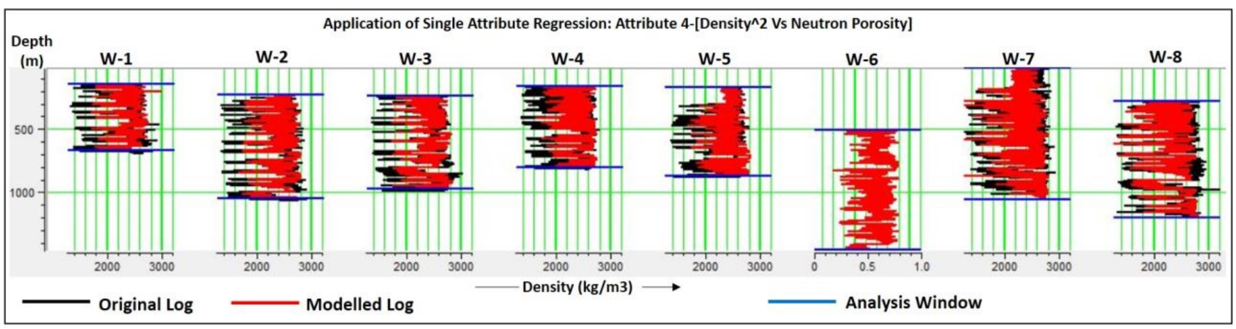

Figure 5: Single Attribute Analysis for the prediction of Density log of well W-6 using Attribute 4-[Density^2 Vs Neutron Porosity].

4. Multi-Attribute Analysis: Stepwise Regression

Multi-attribute analysis has been carried out by creating a list of transform by stepwise regression with suitable operator length, initiating target as obtained from the best attribute of the single attribute analysis result (Table-2). Regression equation are generated by curve fitting technique which uses statistical measure to determine the strength between one dependent and a series of other changing variable. The objective is to derive a multi attribute transform, which is a linear or nonlinear transform between a subset of the attributes and the target log values. A much faster, although less optimal procedure is stepwise regression (*Draper and Smith, 1966*) which is based on the assumption that if the best combination of M attributes is already known, then the best combination of (M + 1) attributes includes the previous M attributes as members with the previously calculated coefficients must be re-derived.

The first procedure in the stepwise regression is to find out single best attribute by exhaustive search by solving the optimal coefficients for the calculation of prediction error. The best Attribute (A) is the one with the lowest prediction error (A_1). Second, the best pair of attributes are found out, assuming that the first member is (A_1). For each other attribute in the list in the form like (A_1, X_1), (A_1, X_2), etc. the optimal coefficients are solved for the prediction error calculation. The best second attribute pair with the lowest prediction error is called as (A_2). Third, the best triplet attributes are found out, assuming that the first two members are A_1 and A_2. For each triplet attribute in the list in the form like (A_1, A_2, Y_1), (A_1, A_2, Y_2), etc. the optimal coefficients are solved and the prediction error are calculated. The best triplet with the lowest prediction error called A_3. This process is carried on as long as desired. The average error of all well Vs attributes is plotted shown in Figure-6A, the histogram plot from the obtained result of predicted and actual density from multi-attribute analysis is shown in Figure-6B which shows the deviation in the predicted value from the observed value of density, and Figure-6C is the cross-plot of the predicted and observed density values obtained from the multi-attribute result of [1/LLS] to generate the regression equation with correlation 0.91 and error 149.2.

S. No.	Target	Attribute	Training Error	Validation Error
1	Density^2	NPHI	214.51	217.38
2	Density^2	Sqrt [GR]	162.09	167.43
3	Density^2	Log [LLD]	149.55	157.14
4	Density^2	1/[LLS]	149.19	154.87

Table 2: List of Multi-Attribute analysis with training and validation error of four attributes transform.

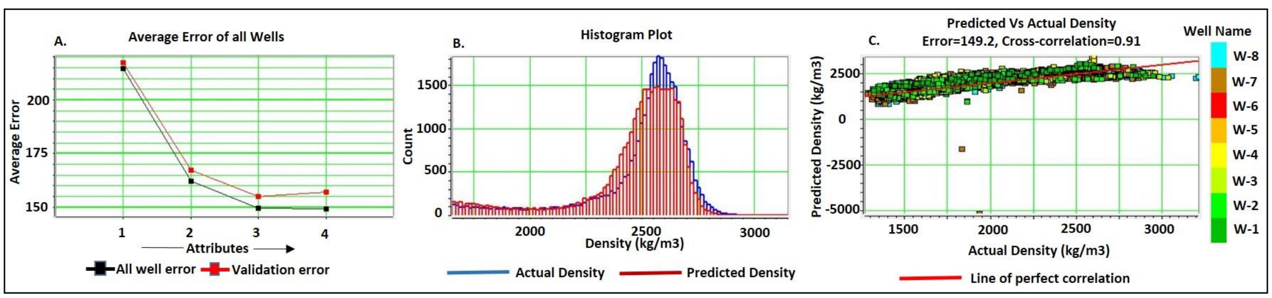

Figure 6: A. Plot of Average error of all wells Vs four multi-attributes attributes. B. Histogram plot of Actual and Predicted density to validate the result. C. Correlation plot of predicted Vs actual density obtained after multi-attribute transform.

5. Probabilistic Neural Network (PNN)

Neural network is a technique used to derive the hidden relationship between the input layers and output targets by applying weights on each layer consisting of nodes that are connected with weights which are determined by the concept of "distance" in multi-dimensional attribute space from known point to the output unknown layer (*Specht, 1990*). Typically artificial Neural Network consists of three layers input, hidden and output, where it is needed to find out the hidden layer (Θ) which is a sigmoid as a transfer function (Figure-7). The PNN is used for classification or mapping by means of cross validation for finding the hidden data on well-by-well basis or point-by-point basis by finding the best sigma (smoothing parameter) value. In mapping technique the weighting functions has been multiplied by the known log value to determine the unknown log value. The output layer has one node as we are predicting a single log property using a single hidden layer, with the number of nodes set by experimentation. The training process is performed by presenting training examples to the network that consists of finding the optimum weights between the nodes. Each example consists of data for a single time sample {A_1, A_2, A_3, L}, where Ai are the attributes and L is the measured target log value. The problem of estimating the weights can be considered a nonlinear optimization problem, where the objective is to minimize the mean-squared error between the actual target and the predicted target log values (*Masters, 1994*). The formula for calculation the least square error is shown below:

$$E=\sum e_i^2 = \sum (d_i^{obs} - d_i^{pre})^2 \qquad …….. \quad (1)$$

Where, d_i^{obs} is the observed data, d_i^{pre} is the predicted data and e_i is the error.

The modelled log after PNN application is shown in Figure-8. The final predicted density log of well W-6 is shown in Figure-9B and Figure-9A is the data before prediction of density log data. The predicted density log is correlated with density log of the nearby wells and observed to have similar lithological trend.

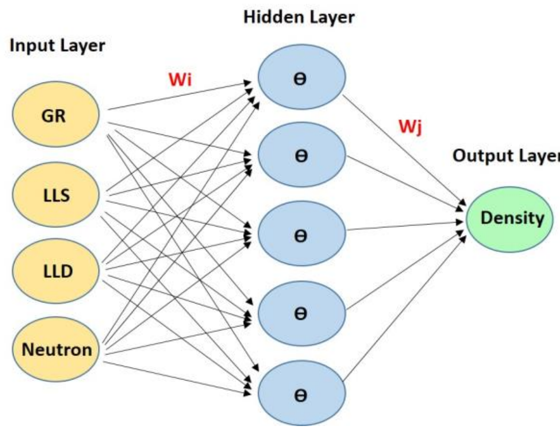

Figure 7: Schematic diagram of Artificial Neural network with Input layer as available parameters, Hidden layer is a sigmoid function represented as Θ and Output layer with single node as density log, used in the methodology and using of weighted function (Wi and Wj) to obtain the output density log curve with adjusted gain.

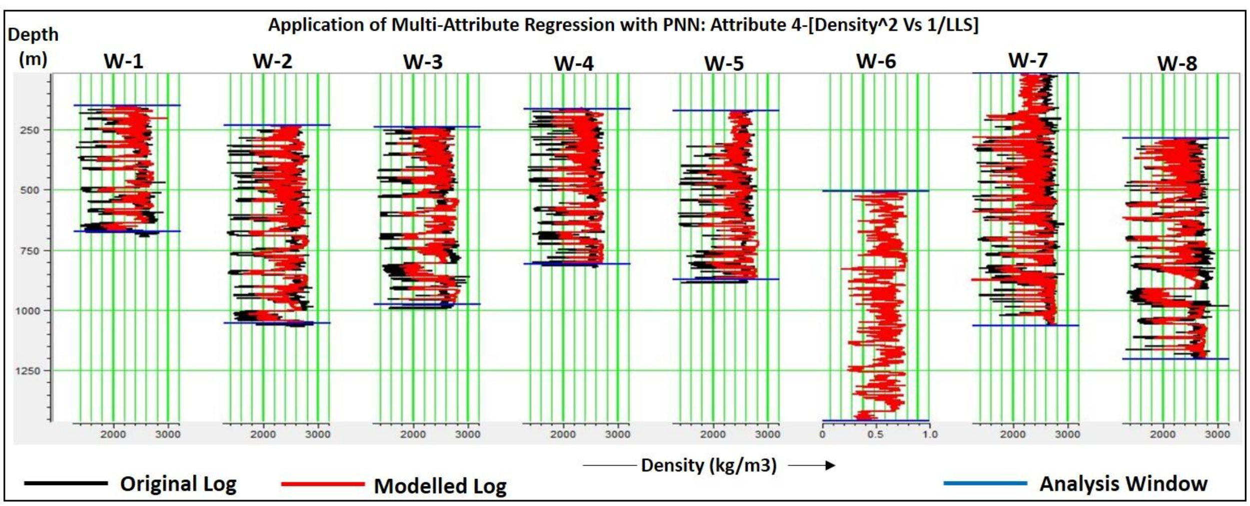

Figure 8: Multi-Attribute analysis and PNN application for the final density log prediction of well W-6 using Attribute 4 [Density^2 Vs 1/LLS].

6. Results and Discussions

The missing log has been predicted in well W-6 using attribute analysis and PNN, the regressing equation has been established in the methodology as discussed in this paper. The correlation result and error obtained by applying multi-attribute analysis using stepwise regression and PNN are 0.91 and 149.2 is better compared to single attribute analysis result with correlation and error as 0.80 and 214.5. The predicted density log of well W-6 has been correlated with the neighboring wells and observed

Application of Attribute Analysis and Probabilistic Neural Network to predict Missing Log data

to have followed similar lithological trend.

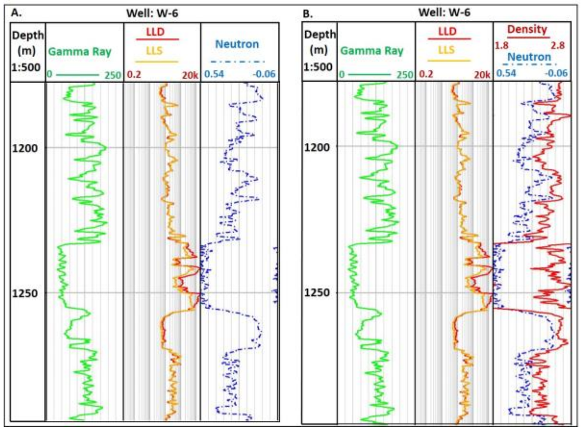

Figure 9: Image representing the composite log (Track-1: Depth, Track-2: Gamma Ray, Track-3: LLD and LLS, Track-4: Neutron) of well W-6, representing the data before (A) and after (B) the prediction of density data (Track-4).

7. Conclusions

The methodology used in this paper is useful in determining the missing log data due to tool malfunctioning or tool failure, which arise beyond the human control. The artificial intelligence technique using attribute analysis and PNN is the future technology to fill the gap between the human and machine error and necessary step for the cost optimization in oil industry with productive output.

References

Atkeson, C. G., 1991, Using locally weighted regression for robot learning. Proceedings of the 1991 IEEE Conference on Robotics and Automation; Los Alamitos, CA: IEEE Computer Society Press, 958-963.

Draper, N. R. and Smith, H., 1966, Applied regression analysis; John Wiley & Sons, Ltd., New York, 407.

Hampson, D. Schuelke, J. and Quirein. J., 2001, Use of multi-attribute transforms to predict log properties from seismic data, Geophysics, 66, 220-236.

Justice, J. H. Hawkins, D. J. and Wong, G., 1985, Multidimensional attribute analysis and pattern recognition for seismic interpretation, Elsevier., 18, 391-399.

Kelly, J.D., Davis. Jr. and L., 1991, Hybridizing the genetic algorithm and the nearest neighbors classification algorithm, Proceedings of the Fourth International Conference on Genetic Algorithms, San Mateo, CA: Morgan Kaufmann., 377-383.

Masters, T., 1994, Signal and image processing with neural networks: John Wiley & Sons, 417.

Specht, D.F., 1990, Probabilistic neural networks. Neural Networks, 3, 1, 109-118.

Acknowledgement

Authors express their gratitude to Shri. R.P. Pandey, Asset Manager, CBM Asset, ONGC Bokaro, for his encouragement and permission to present this paper. Authors also wishes to place on record the help and guidance rendered by Department of Applied Geophysics, IIT-ISM, Dhanbad, Sub-Surface Team and Logging Services, CBM Asset, ONGC Bokaro for this publication.