APPLICATION OF DEEP LEARNING TECHNIQUES FOR LITHOFACIES CLASSIFICATION USING WELL LOG DATA OF TARANAKI BASIN

Mayank Verma, ONGC; Dr. Ravi Sharma, Assistant Professor, Dept of Earth Sciences, IIT Roorkee

verma_mayank1@ongc.co.in

Keywords

Lithofacies Characterization, Automation, Machine Learning, SVM, Neural Networks

Summary

Classification of different lithofacies is crucial in seismic interpretation because different rocks have different permeability and fluid saturation for a given porosity. The ideal sources for lithofacies classification are core samples of rocks extracted from wells. Nevertheless, core samples cannot always be obtained due to associated costs. The conventional classification method is based on manually assigning lithofacies by human interpreters and is a very tedious and time-consuming process. We aimed at automating this classification process through the use of machine learning and deep learning methods. We had selected wells from the region which have the mud log and wireline logs present with them. Machine learning algorithm, Support vector machines (SVM) was employed to build an automatic lithofacies classifier and accuracy of the model was validated on a well with unlabelled lithofacies and accuracy of 0.63 was achieved. In the next phase, a Deep learning (DL) model based on Convolutional Neural Network was developed and it achieved a classification accuracy of 0.71. Due to the skewed nature of the dataset, the validation accuracy of the model showed a stark drop when compared with training accuracy. This major drop in accuracy occurs while classifying those facies which have limited number of training example. The accuracy could be further improved by incorporating adjacent lithofacies in classification task, which was the limitation of the target dataset.

Introduction

Facies, in geology, is basically a way to differentiate rock bodies into units which are mappable, in terms of composition, characteristics, formation or several other attributes. Facies are used by geologists to group together body of rocks with similar characteristics in order to facilitate the study of a basin of interest. In the case of Oil & Gas reservoirs, porosity and permeability are critical properties to determine since they give indications about the potential volume of fluids that might be stored in a rock and how they will flow during production. We can therefore expect that grains size, shape and density as well as the depositional and compaction history of the rocks will be a dominant factor for the categorization.

While the main source of information for defining those facies comes from the observation of core samples under visible and x-ray light, we also have a variety of well log recordings at our disposal. By measuring the acoustic and electrical responses as well as the nuclear radiations of the drilled medium, we can infer properties about its rock matrix and fluid content and indirectly relate them to the porosity, permeability or fluid saturation of the rocks.

A number of quantitative and qualitative methods for lithology interpretation are already employed that combine various metrics.

Qualitatively, we inefficiently evaluate measurements from logging operations along with Photo electric (Pe) factor analyses, gamma ray evaluations done for identifying shale (Gardner and Dumanoir (1980), Serra et al. (1985), Dewan (1986), or using multiple logs. This proves insufficient as several lithologies pose complexity which requires larger set of information than just provided through these.

Within time, lithology interpretation has expanded and starts to consider the usage of quantitative methods, such as Crossplots, statistical analysis, and neural network.

Burke et al. (1969) was the first one to introduce Crossplots based quantitative analysis which is commonly used over neutron, Pe, density and sonic. It involves simultaneous plots of two or more log data points. Even Clavier and Rust (1976) studied its application in quantitative analyses. These methods have a limitation that they require manual human intervention and cannot be automated for the task.

Statistical Analysis, proven to have applications for lithological identification by Delfiner et al. (1987), laid way for combining wireline measurement done these days, to generate automated lithological description.

Busch et al. (1987) demonstrated lithological prediction was possible using statistical analysis. It proved the importance of Bayesian rule for probability and discriminant analysis in lithological classification. But, this method has a limitation in the criteria for the data to be of normal distribution based (limits the geophysical dataset).

We aimed at automating this facies classification process through a method of indirect quantification (e.g., wireline logs.). Machine learning algorithm, Support vector machines (SVM) & Deep learning model based on Convolutional Neural Network was employed to build an automatic lithofacies classifier.

Data and Methodology

The dataset is from Taranaki basin which is situated on the west coast of New Zealand. It is an onshore-offshore rift basin. The target wells selected for study were from KAURI region of the Taranaki basin.

The ideal log for the purpose of lithological classification should be: Most affected by rock properties & Least affected by fluid properties

The available logs in the Kauri regions based on above criteria were: Caliper (CALI), Gamma ray (GR), Formation density (DENS), Photoelectric absorption (PEF), Neutron porosity (NEUT), Sonic log (DTC), Resistivity log deep (RESD).

In the first step, the selection of 8 wells from Kauri region was done out of nearly 600 available wells having appropriate wireline logs and mud log information. Here the dataset has 8 lithofacies type and the label will be assigned to each facies type in order to use it as a predicted variable. METAMORPHIC (MMP), CLAYSTONE (CST), SANDSTONE (SST), METASEDIMENT (Msdt), SILTSTONE (SiS), CONGLOMERATE (CONG), LIMESTONE (LST), COAL (COAL).

Out of these 8, seven wells were used for training purpose and 1 well was used for validation purpose (Fig.1)

In SVM: Using a technique known as Kernel trick, the data is projected to higher dimensional space where it is separable into distinct classes.

In CNN: Build a classifier using the image classification technology for geoscientific problems.

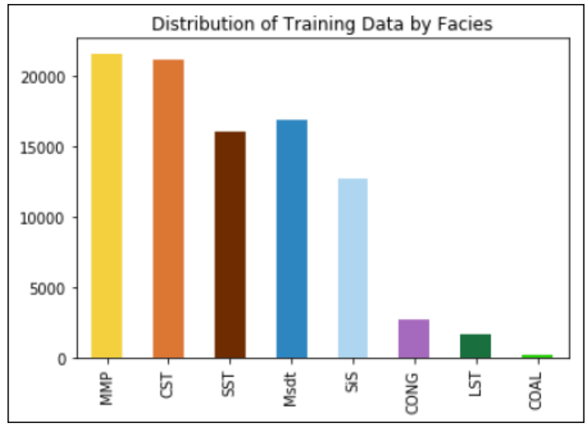

Figure 1: shows the distribution of different facies in the training dataset. It can be easily inferred by visual analysis that the data is skewed and have very few examples of COAL lithology as compared with the other lithofacies.

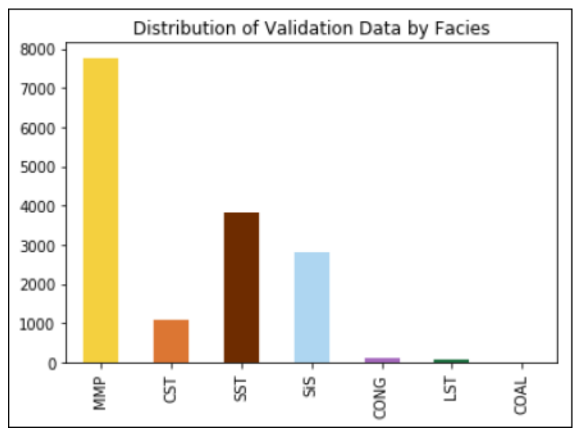

Figure 2: shows the distribution of different facies in the validation dataset. This dataset has ZERO example of Metasediment (Msdt) and very large number of Metamorphic (MMP) facies example. This type of distribution very heavily affects the prediction ability of the classification model.

Results

Hyper parameter tuning in SVM - The model has trained a series of classifiers with different values for C (hypermeter to control error) and gamma (hypermeter to give curvature weight of the decision boundary). The classification accuracy is recorded for each combination of parameter values (Fig. 3).

GAMMA	C	Training_Accuracy	Validation_Accuracy
1	1	0.89	0.58
1	10	0.81	0.63
10	1	0.89	0.58
10	10	0.89	0.58

Figure 3: Hyper parameter selection table for SVM

Hyper parameter tuning in CNN - The model has trained a series of classifiers with different values for Drop out (regularization technique where randomly selected neurons are ignored during training) and k-Fold (split the input data into k subsets of data, train on (k-1) subsets and evaluate on subset that was not used for training).

Drop out	Fold	Training_Accuracy	Validation_Accuracy
0.3	6	0.697	0.70
0.4	6	0.713	0.70
0.3	3	0.704	0.71
0.6	3	0.701	0.67

Figure 4: Hyper parameter selection table for CNN

The seven plots show the measured logs as a function of depth (in meters). The two facies series represent the prediction by the machine on the right and the solution proposed by the geologists on the left (Fig. 5b & Fig. 6b)

The F1 score combine both accuracy and precision to give a single measure of relevancy of the classifier results.
Our SVM classifier achieved an overall F1 score of 0.63 on the test well, so there is room for improvement.

	precision	recall	f1-score	support
MMP	0.98	0.92	0.95	7771
CST	0.14	0.07	0.10	1080
SST	0.50	0.41	0.45	3824
Msdt	0.00	0.00	0.00	0
SiS	0.36	0.36	0.36	2804
CONG	0.07	0.64	0.12	115
LST	0.04	0.12	0.06	83
COAL	0.00	0.00	0.00	15
avg / total	0.68	0.63	0.65	15692

Figure 5a: dependency of precision and recall probabilities on the support. Higher the support of a particular facies, higher the f1-score i.e. higher probability of classification model to correctly classify it.

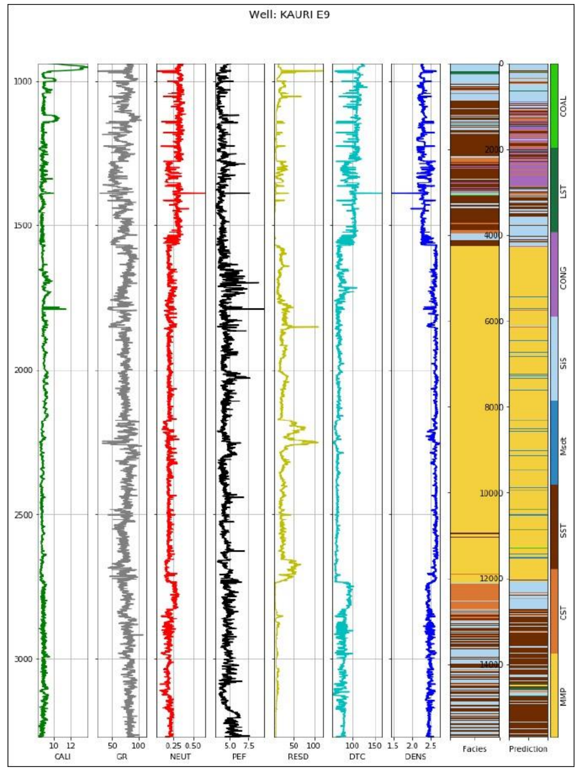

Figure 5b: shows the result of classification with the model parameters Gamma = 1 and C = 10. The classifier achieved an accuracy of 0.63 in classifying data from well using SVM classifier (KAURI E9)

The CNN classifier achieved an overall F1 score o 0.71 on the test well, so CNN performed better than SVM.

	precision	recall	f1-score	support
MMP	0.97	0.96	0.97	7771
CST	0.01	0.01	0.01	1080
SST	0.72	0.41	0.52	3824
Msdt	0.00	0.00	0.00	0
SiS	0.43	0.72	0.54	2804
CONG	0.13	0.46	0.20	115
LST	0.20	0.01	0.02	83
COAL	0.00	0.00	0.00	15
avg / total	0.74	0.71	0.70	15692

Figure 6a: high dependency of model on the number of training examples. A positive correlation is found between the f1-score and support of different facies.

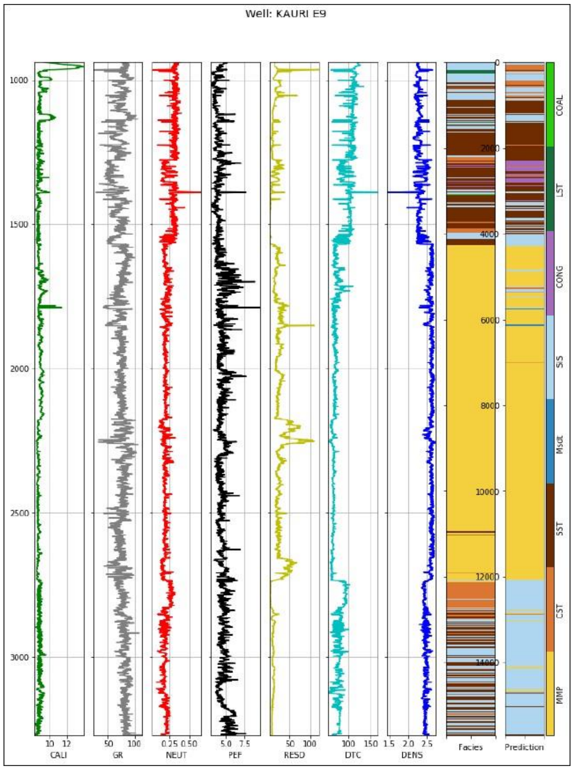

Figure 6b: shows the result of classification with the model parameters Drop out = 0.3 and Fold = 3. The classifier achieved an accuracy of 0.71 in classifying data from well using CNN classifier (KAURI E9)

Conclusions

From the work carried out, it can be concluded that machine learning and deep learning techniques can be applied to predict the lithofacies of the wells in the regions where we have wireline logs for all the wells and only few wells with the mud logs available (to train the classifiers) with us.

Results obtained on a set of seven wells validate the proposed approach, which highlights the positive impact of the developed feature augmentation strategy.

The results obtained while validating our model on blind well also give a confirmation to a good capacity of this model to generalize to new data. Using deep learning strategies for feature learning and classification (e.g., ConvNets), improved results were obtained.

Both methods have limitation in terms of skewed dataset therefore resulted in poor precision and recall score for those particular facies.

Both these methods have very high dependency to the support of different lithofacies in the training samples. Higher the support of a particular facies in the training dataset, higher is the probability of correctly classifying that particular facies to its correct class.

Considering the achieved promising results, future work will be devoted on validating the possibility of adding geological constraints to drive classification with the help of apriori information about rock formation.

References

Gardner, J. S. and Dumanoir, J. (1980). Litho-Density Log Interpretation. In SPWLA1980-N, SPWLA. Society of Petrophysicists and Well-Log Analysts.

Serra, O., Delfiner, P., and Levert, J. C. (1985). Lithology Determination From Well-Logs: Case Studies. In SPWLA-1985-WW, SPWLA. Society of Petrophysicists and Well-Log Analysts.

Dewan, J. T. (1986). Open-Hole Nuclear Logging - State Of The Art. In SPWLA-1986-MM, SPWLA. Society of Petrophysicists and Well-Log Analysts.

Burke, J., Campbell, Jr., R., and Schmidt, A. (1969). The Litho Porosity Cross Plot: A New Concept For Determining Porosity And Lithology From Logging Methods. In SPWLA-1969-Y, SPWLA. Society of Petrophysicists and Well-Log Analysts.

Clavier, C. and Rust, D. (1976). Mid Plot:a New Lithology Technique. SPWLA-1976-vXVIIn6a2.