Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Comparison of Machine Learning Models for Porosity Prediction

Pedro Baptista[1] Thomás J. Spier[2]
Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, MG
Henrique Versieux[3]
Departamento de Matemática, Universidade Federal de Minas Gerais, Belo Horizonte, MG
Caique Carvalho[4] Maria de Oliveira[5]
CPMTC-IGC, Instituto de Geociências, Departamento de Geologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG

In petroleum exploration research, a primary objective is to efficiently assess whether a petroleum reservoir exists in a particular region using minimal resources. Given the significant costs and labor involved in drilling exploration wells, it is imperative to seek ways to minimize these efforts by predicting as much as possible about the properties of sediment layers and reservoirs in the target area.

Seismic inversion is a technique used in geophysics and petroleum exploration to estimate the properties of subsurface rock formations based on seismic data. Seismic data are collected by sending acoustic waves into the earth and recording the reflections that bounce back. These reflections provide information about the various layers of rock underground.

Seismic reflections can be modelled by the convolution of the reflection coefficients and the seismic wavelet as follows:

$$d(t,\theta) = w(t,\theta) * r(t,\theta) \tag{1}$$

where $d(t,\theta)$ is the reflection seismic response, $w(t,\theta)$ is the seismic wavelet, meaning, the energy pulse that is sent into the subsurface during data acquisition and $r(t,\theta)$ is the reflection coefficients, i.e., the coefficients that indicate the magnitude of the energy being reflected in relation to the incident energy, and it varies according to the elastic properties of the different layers in the subsurface. This model is based on the time variable of the wave propagation and the incidence angle.

Usually, the first step in the seismic inversion is to perform an inversion to obtain the elastic properties of the rock layers. Once a good estimate of these rock properties is obtained, a second step is performed in order to obtain rock physics models, allowing to recover information such as porosity or fluid saturation from elastic properties; see for instance [1] and references therein.

In this work, we study rock physics models to estimate porosity based on elastic properties: velocities v_p and v_s and density ρ of the different geological layers. The models are derived from a dataset provided by ANP (Brazilian National Agency of Petroleum) of real wells from Buzios oil field.

[1] pedro.baptista@dcc.ufmg.br
[2] thomasjung@dcc.ufmg.br
[3] hversieux@ufmg.br
[4] caiquepc@ufmg.br
[5] mariaoliveira.tl@gmail.com

Four different rock physics models to predict porosity are considered in our work. First, a simple linear regression. Next, two models are derived using a similar strategy: initially a clustering method, either K-Means or Gaussian Mixtures Model is used to separate v_p, v_s and ρ data into a different number of clusters, then a linear regression is applied in each cluster. Finally, we use a supervised machine learning method called Light Gradient Boosting Machine (LightGBM, for short); [2]. The LightGBM is based on decision trees and is proposed to handle datasets with large scale. It uses novel techniques to reduce the feature's dimension in order to increase its efficiency, and has been previously applied to predict porosity [3].

We present a comparison of the different methods using a cross validation test. More precisely, the models are trained on subsets of the wells' data, while one well is used only to test the corresponding derived model. Prediction errors are calculated considering different combinations of training and testing wells.

Additionally, to evaluate the robustness of our methods, we introduce varying levels of noise into the test data. We conduct comparisons by replicating a well and introducing different independent noise sources, thereby expanding our sample for comparison.

Quantitative and qualitative assessments are employed for comparison. Quantitatively, we use the mean squared error measure, while qualitatively, we examine graphs. It's worth noting that a method could potentially reduce the error metric by converging to a simple average of the expected porosity.

Our preliminary findings, based on a subset of five wells, indicate that both the K-means and LightGBM algorithms perform better than the other models in the noiseless experiments. However, as we introduce Gaussian noise with increasing amplitudes, LightGBM demonstrates a more resilient inversion capability, followed by simple linear regression as the next best performer. The Gaussian mixture model consistently yields the poorest results across all noise levels.

Qualitatively, all inversion methods seem to approximate the curves well, rather than aiming for a mean approximation.

Acknowledgements

PB, CC, MO, TS and HV acknowledge the support of CNPq. All the authors acknowledge the support from Petrobras through PetrobrasIAGeo project. All the authors acknowledge ANP for providing the dataset used.

References

[1] D. Grana, T. Mukerji, and P. Doyen. **Seismic reservoir modeling: Theory, examples, and algorithms**. John Wiley & Sons, 2021. ISBN: 9781119086185.

[2] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu. "LightGBM: a highly efficient gradient boosting decision tree". In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 3149–3157. ISBN: 9781510860964.

[3] C. Zou, L. Zhao, F. Hong, Y. Wang, Y. Chen, and J. Geng. "A comparison of machine learning methods to predict porosity in carbonate reservoirs from seismic-derived elastic properties". In: **Geophysics** 88.2 (Mar. 2023), B101–B120. ISSN: 0016-8033. DOI: 10.1190/geo2021-0342.1.

