Fracture Conductivity Prediction Based on Machine Learning in Shale – Decision Tree and Random Forest Regression

Ella Meyer*

*Address all correspondence to: Ella Meyer, E-mail: ella.meyer22od2@gmail.com

KEYWORDS: Fracture conductivity, Machine Learning, Shale, Decision Tree, Random Forest

Abstract: Hydraulic fracturing extracts oil and gas from deep underground, with fracture conductivity being crucial for efficient production. Traditional lab techniques for measuring conductivity are costly and time-consuming. This paper explores using machine learning, specifically decision tree and random forest regression, to predict fracture conductivity based on experimental data like Poisson's ratio and proppant size. Optimizing these models can enhance hydraulic fracturing efficiency in shale formations.

Hydraulic fracturing, often referred to as fracking, is a technique used to extract oil and gas from deep underground formations. This process involves injecting a high-pressure fluid mixture into the rock layer, creating fractures through which hydrocarbons can flow more freely to the production well. A crucial aspect of hydraulic fracturing is fracture conductivity, which refers to the ability of the created fractures to allow the flow of hydrocarbons. High fracture conductivity is essential for maintaining efficient hydrocarbon production. It depends on various factors, including the proppant type and size, the closure stress, and the properties of the fracturing fluid. Optimizing these factors ensures that the fractures remain open and conductive over the lifespan of the well (Montgomery, 2013; Holditch, 2006; Sharma et al., 2014).

Determining fracture conductivity in situ is a very challenging task and conducting lab measurements using Hassler core technique on shale are expensive and time consuming. The Hassler core experiment is a laboratory technique that simulates in-situ conditions to measure the permeability and conductivity of fractures in core samples. This method involves placing a core sample within a Hassler sleeve, applying confining pressure, and injecting fluid to simulate the hydraulic fracturing process. Researchers measure the resultant fracture conductivity, which is a key parameter influencing the efficiency of hydrocarbon extraction from shale formations. (Wu et al., 2017, Wu et al., 2019).

Machine learning is being used to advance scientific computing in a variety of fields, such as fluid mechanics, (Raissi et al., 2019, Zhang et al., 2020, Wang el al., 2017) solid mechanics, (Haghighat et al., 2020, Zheng et al., 2022, Arora et al., 2022) materials science etc (Kim et al., 2021). Machine learning techniques can also be applied to predict the fracture conductivity of new shale formations based on models built using data collected from experiments.

Good experimental data is key to building a well-performing prediction model. In order to build a model to predict fracture conductivity, the following information about various shale samples needs to be recorded while conducting experiments: Poisson's ratio, Young's Modulus, Temperature, Closure pressure, Proppant particle size, and Sand concentration. The wider the range of the above-mentioned parameters, the better the predictions will be across different shale formations.

There are various machine learning techniques that can be used to build the above model. In this paper we will discuss about decision tree and random forest regression.

Data Preprocessing

Once the data collection phase is complete, we need to preprocess the data. This includes handling any missing values, normalizing or standardizing the data to ensure that all parameters are on a comparable scale, and possibly performing feature selection to identify the most significant predictors of fracture conductivity. Preprocessing is crucial as it directly impacts the performance of both the decision tree and random forest models, helping to eliminate biases and improving the overall quality of the dataset.

Decision Tree Regression

A decision tree regression model predicts the target variable by learning simple decision rules inferred from the data features. The process starts by splitting the dataset into subsets based on the feature that results in the most significant reduction in variance. The splits are made recursively, creating a tree structure where each node represents a decision based on a single feature, and each leaf node represents a predicted value. Mathematically, the variance reduction or information gain for a feature split can be represented as:

$$\Delta \mathrm{Var}(S, f) = \mathrm{Var}(S) - \left(\frac{|S_l|}{|S|} \mathrm{Var}(S_l) + \frac{|S_r|}{|S|} \mathrm{Var}(S_r) \right)$$

where S is the set of data points, S_l and S_r are the subsets after the split, and f is the feature used for the split.

Random Forest Regression

A random forest regression model is an ensemble learning method that constructs multiple decision trees during training and outputs the average of their predictions. This approach helps to mitigate the overfitting problem commonly associated with individual decision trees. Each tree in the forest is trained on a different bootstrap sample of the data, and a random subset of features is considered for splitting at each node. The final prediction is obtained by averaging the predictions from all the trees:

Model Evaluation and Validation

After training the decision tree and random forest models, their performance is evaluated using metrics such as the coefficient of determination (R^2), mean squared error (MSE), and root mean squared error (RMSE). Additionally, cross-validation techniques, such as k-fold cross-validation, are used to ensure the models' reliability and to avoid overfitting. For example, in k-fold cross-validation, the data is divided into k subsets, and the models are trained k times, each time using a different subset as the validation set and the remaining $k-1$ subsets as the training set. The performance metrics are then averaged to obtain a more robust estimate of the models' performance.

Conclusion

By following these steps, comprehensive and accurate decision tree and random forest models can be developed, providing valuable insights for predicting the fracture conductivity of shale formations. The combination of good experimental data, thorough preprocessing, appropriate model selection, and rigorous validation ensures the models' reliability and applicability in real-world scenarios. These models not only

predict fracture conductivity but also help in understanding the importance of different features in influencing the predictions.

References

Montgomery, C. T. (2013). "Fracturing Fluids." In Hydraulic Fracturing (pp. 67-98). Elsevier.

Holditch, S. A. (2006). "Tight gas sands." Journal of Petroleum Technology, 58(06), 86-93.

Sharma M. M et al. (2014). "Proppant Selection and Its Effect on Fracture Conductivity and Production." SPE Hydraulic Fracturing Technology Conference.

Wu W, Kakkar P, Zhou J. et al. 2017. An Experimental Investigation of the Conductivity of Unpropped Fractures in Shales. Presented at the SPE Hydraulic Fracturing Technology Conference and Exhibition, The Woodlands, Texas, 24-26 January. SPE-184858-MS. https://doi.org/10.2118/184858-MS.

Wu W, Zhou J, Kakkar P et al. An experimental study on conductivity of unpropped fractures in preserved shales. SPE Production & Operations, 2018, 34(2): 280–296

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378: 686–707, 2019. DOI: https://doi.org/10.1016/j.jcp.2018.10.045.

T. Zhang, B. Dey, P. Kakkar, A Dasgupta, and A. Chakraborty. Frequency-compensated pinns for fluid-dynamic design problems. arXiv preprint arXiv:2011.01456, 2020. DOI: https://doi.org/10.48550/arXiv.2011.01456.

J. Wang, J. Wu, and H. Xiao. Physics-informed machine learning approach for reconstructing reynolds stress modeling discrepancies based on dns data. Physical Review Fluids, 2(3):034603, 2017. DOI: https://doi.org/10.1103/PhysRevFluids.2.034603.

E. Haghighat, M. Raissi, A. Moure, H. Gomez, and R. Juanes. A deep learning framework for solution and discovery in solid mechanics: linear elasticity. arXiv preprint arXiv:2003.02751, 2020.

B. Zheng, T. Li, H. Qi, L. Gao, X. Liu, and L. Yuan. Physics-informed machine learning model for computational fracture 41 of quasi-brittle materials without labelled data. International Journal of Mechanical Sciences, 223:107282, 2022. DOI: https://doi.org/10.1016/j.ijmecsci.2022.107282.

R. Arora, P. Kakkar, B. Dey, and A. Chakraborty. Physics-informed neural networks for modeling rate- and temperature dependent plasticity. arXiv preprint arXiv:2201.08363, 2022. DOI: https://doi.org/10.48550/arXiv.2201.08363

Y. Kim, C. Yang, K. Park, G. X. Gu, and R. Seunghwa. Deep learning framework for material design space exploration using active transfer learning and data augmentation. npj Computational Materials, 7:140, 2021. DOI: https://doi.org/10.1038/s41524-021-00609-2

E. Meyer. Fracture Conductivity Prediction Based on Machine Learning in Shale – Part I, 2024 DOI: https://doi.org/10.31224/3757