Machine learning as a tool to predict the mass of oil from well logs

Marcelo Guarido and Daniel Trad
CREWES – University of Calgary

Summary

Oil saturation is the measure of the amount of oil inside the porosity of a reservoir rock. Its calculation, usually from core analysis, is an important quantity that helps to characterize the reservoir. In this work, we are not predicting the actual oil saturation due to the lack of information for the wells gathered, but the fraction of the mass of oil in the core. Most of this work is focused on the data preparation before modelling, as our variables and targets came from two different measurement sources (well logs and core analysis), and in how to create a valid workflow to make features and targets compatible with each other. In the end, we show how to select an appropriate machine learning model to predict the target, which needs to be one with non-linear properties, and how to interpret the feature importance. To predict the fraction of mass of oil, the induction log ILD is the one that brings most of the information, but it needs to be combined with other logs for the prediction to make sense. The metric used to evaluate the models was the R^2, and the best model had a score of 0.82.

Introduction

Knowledge of oil and water saturation distribution in an oilfield is of great significance for the later development (Yue et al. 2018), and direct petrophysical models can do the estimation (with fair knowledge of the relations between logs and saturation) or, more recently, with the use of machine learning models (Zhang et al. 2019). Here we will focus on the machine learning side of it.

Machine learning algorithms usage on geoscience, engineering, and petrophysics data is in ascension this last decade. Maybe the most common application is for facies classification, by the use of ensemble classifiers (Bestagini, Lipari, and Tubaro 2017; Zhang and Zhan 2017; Caté et al. 2017), neural networks (Silva et al. 2014), and support vector machines (Caté et al. 2017; Alexsandro, P. Carlos, and Geraldo 2017; Wrona et al. 2018). Guarido (2019) showed that deeply analyzing the data before modeling can provide insights for data engineering, and he applied polar coordinates transformation of the features by realizing a circular relationship with the target. Machine learning has broad applications in geophysics, such as in FWI, by using convolutional neural networks for salt identification (Lewis and Vigh 2017; Guarido, Li, and Cova 2018), and FLEXWIN for time-window selection (Chen et al. 2017). It is possible to find works on trace interpolation using support-vector regression (Jia and Ma 2017), or by Monte-Carlo approximations (Jia, Yu, and Ma 2018). Deep neural networks are used by Araya-Polo et al. (2017) for fault detection and by Araya-Polo et al. (2018) for tomography. Nearest neighbors (k-NN) can be implemented to help with CMP velocity analysis (Smith 2017). Russell, Ross, and Lines (2002) combine NN with AVO. In petrophysics, Ahmadi and Chen (2019) show that hybrid methods provide higher accuracy than single models, but the latter is more robust. Many other machine learning algorithms applications can be found in the literature.

Oil and water saturations also have a fairly long list of machine learning applications to help in interpretation, but we will narrow the list a few. Zhang et al. (2019) use RNN (recurrent neural networks), most specifically the LSTM (long-short term memory) approach to estimate water saturation with good performance. Khan, Tariq, and Abdulraheem (2018) also goes on the deep learning side by training a neural network model to predict water saturation in complex lithologies with high accuracy. Kapoor (2017) shows that ensemble tree methods, such as random forest and gradient boosting, can predict oil and water saturation in the oil sands more accurately than empirical estimations.

In this work, we were not able to work on the estimation of oil saturation in the oil field, as this information is missing on public data (even with the header pointing that it exists). But we were able to retrieve the fraction of mass of oil presented in the core analysis for several wells, and the goal became to estimate it from the well logs.

Data Preparation and Modeling

In this project we analyzed, processed, and modelled wireline logs and core samples from 50 wells at the Athabasca oil field. Figure 1 is an example of the well logs and the core parameter we want to predict: the fraction of mass of oil.
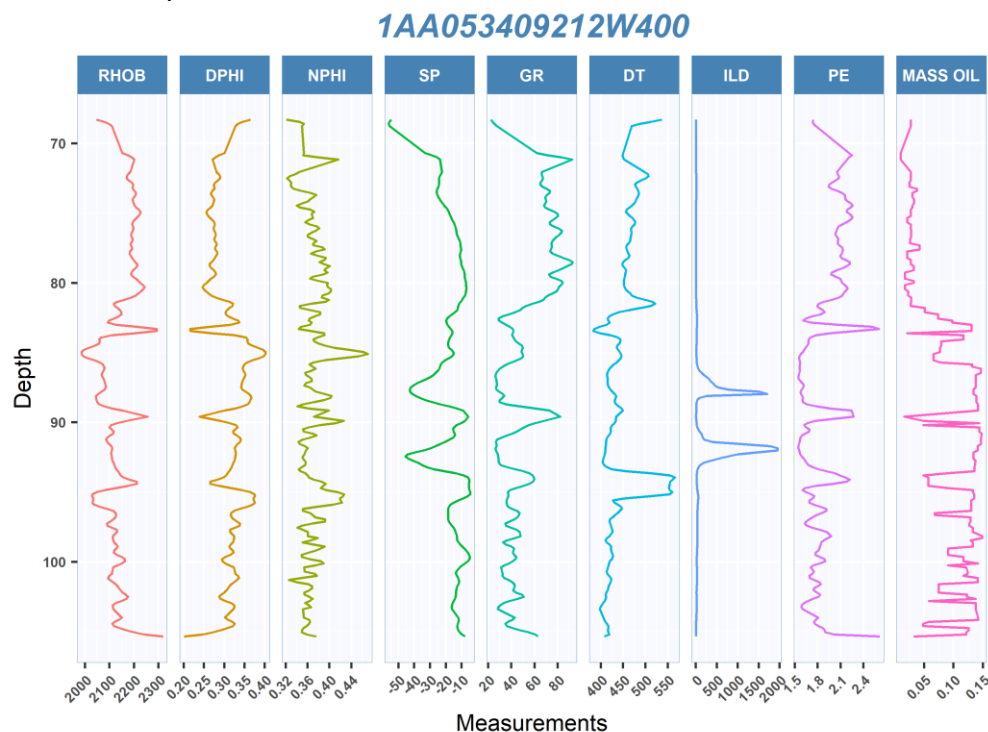

Figure 1: Well logs and core analysis (mass of oil) of one of the 50 wells.

Severe data cleaning and imputation took most of our time during this project, but one of the most important steps prior to modeling was to align the core measurements to the wireline logs. For that, we used a "common variable" from the two sources of data, the porosity. Core analysis has the porosity at a specific level calculated, while the wireline logs came with the "density porosity" DPHI. The measures at not exactly equal, but are compatible, and Figure 2

shows the result of the shift needed to match both traces. The shifts are calculated for each well individually and applied to the core data.
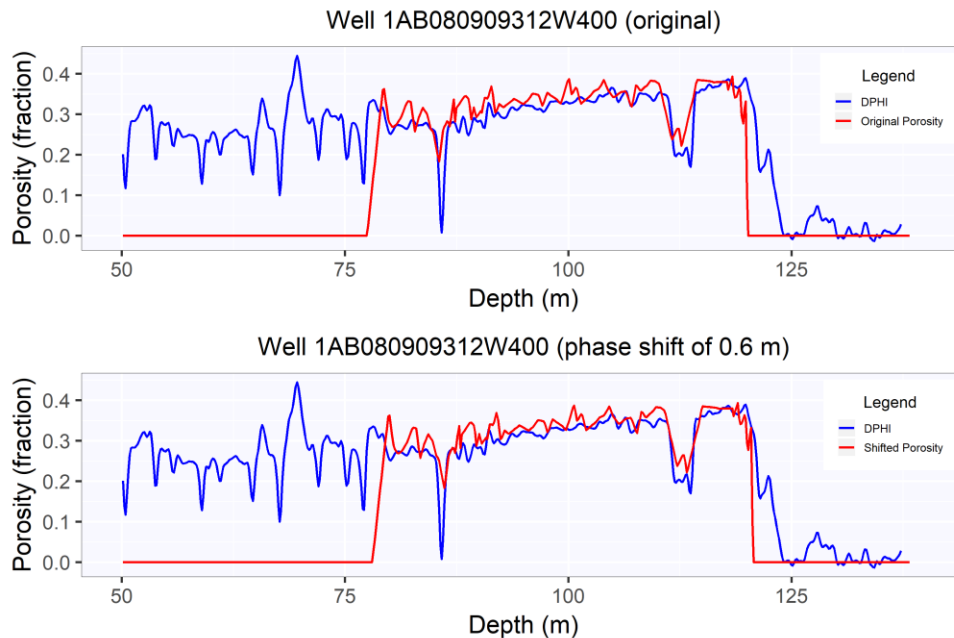

Figure 2: Matching the core porisity to the density porosity.

With all the data matched, we were able to do the modeling and evaluation of the trained model. For this process, we separated 10 wells to be used as the validation set, and the remaining 40 wells are now the training set. This is a regression problem, and the method we selected to predict the mass of oil was the *gradient boosting regressor* (Guarido 2018).
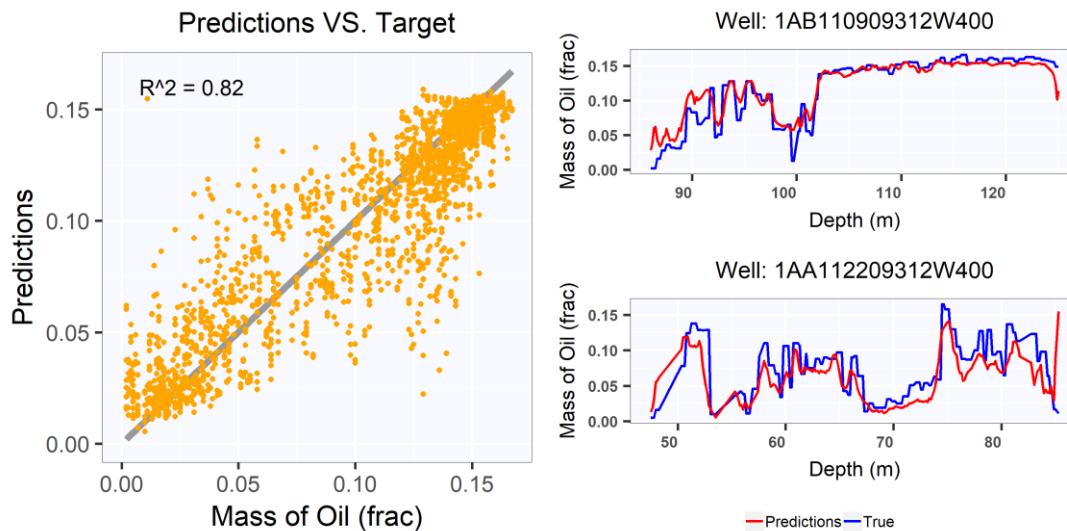

Figure 3: Predicted fraction of mass of oil for the 10 validation wells.

Figure 3 shows the cross-plot of the predicted and true values of the fraction of mass of oil for the 10 validation wells (orange dots), were the gray line is the optimal prediction (how it

should be if the predictions and true values matched 100%), and the R^2 is 0.82. The same figure shows the comparison of the true and predicted values for two wells, showing a close match.
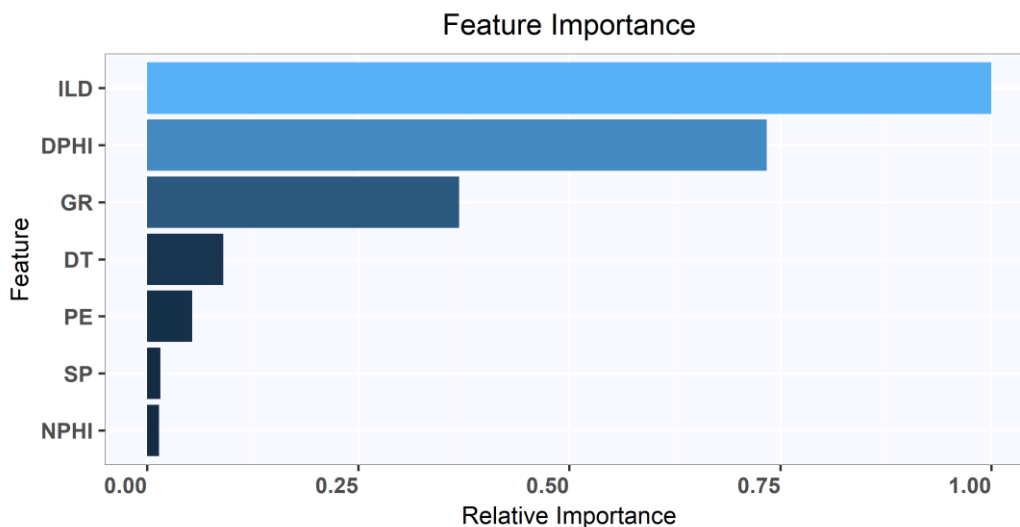

Figure 4: Relative feature importance for the gradient boosting model.

Figure 4 shows the relative importance of the features used for the modeling, were the induction, density porosity, and gamma-ray logs presented to be the best indicator to predict the fraction of mass of oil.

Conclusions

In this work, we were able to show the relevance of machine learning methods to help interpreters in their jobs by predicting the fraction of mass of oil in a core sample using well logs. Data preparation was a very important step for the project, as the data required severe cleaning and processing. The predicted fraction of mass of oil with the gradient boosting regressor showed to be robust and presented a close match to the true values at the validation wells. Also we could check the high importance of the induction, density porosity, and gamma-ray logs to predict the fraction of mass of oil.

Acknowledgments

References

Ahmadi, Mohammad Ali, and Zhangxing Chen. 2019. "Comparison of Machine Learning Methods for Estimating Permeability and Porosity of Oil Reservoirs via Petro-Physical Logs." Petroleum 5 (3): 271–84.

Alexsandro, G. C., A. C. da P. Carlos, and G. N. Geraldo. 2017. "Facies Classification in Well Logs of the Namorado Oilfield Using Support Vector Machine Algorithm." In 1853–8. 15th International Congress of the Brazilian Geophysical Society EXPOGEF, Rio de Janeiro, Brazil, 31 July-3 August 2017.

Araya-Polo, Mauricio, Taylor Dahlke, Charlie Frogner, Chiyuan Zhang, Tomaso Poggio, and Detlef Hohl. 2017. "Automated Fault Detection Without Seismic Processing." The Leading Edge 36 (3): 208–14.

Araya-Polo, Mauricio, Joseph Jennings, Amir Adler, and Taylor Dahlke. 2018. "Deep-Learning Tomography." The Leading Edge 37 (1): 58–66.

Bestagini, Paolo, Vincenzo Lipari, and Stefano Tubaro. 2017. "A Machine Learning Approach to Facies Classification Using Well Logs." In 2137–42. SEG Technical Program Expanded Abstracts 2017.

Caté, Antoine, Lorenzo Perozzi, Erwan Gloaguen, and Martin Blouin. 2017. "Machine Learning as a Tool for Geologists." The Leading Edge 36 (3): 215–19.

Chen, Yangkang, Judith Hill, Wenjie Lei, Matthieu Lefebvre, Jeroen Tromp, Ebru Bozdag, and Dimitri Komatitsch. 2017. "Automated Time-Window Selection Based on Machine Learning for Full-Waveform Inversion." In 1604–9. SEG Technical Program Expanded Abstracts 2017.

Guarido, Marcelo. 2018. "Machine Learning in Geoscience: Facies Classification with Features Engineering, clustering, and Gradient Boosting Trees." CREWES Research Report 30: 13.1–13.23.

Guarido, Marcelo. 2019. "Machine Learning Strategies to Perform Facies Classification." GeoConvention 2019 Abstracts.

Guarido, Marcelo, Junxiao Li, and Raúl Cova. 2018. "Machine Learning in Geoscience: Using Deep Learning to Solve the TGS Salt Identification Challenge." CREWES Research Report 30: 14.1–14.12.

Jia, Yongna, and Jianwei Ma. 2017. "What Can Machine Learning Do for Seismic Data Processing? An Interpolation Application." Geophysics 82 (3): V163–V177.

Jia, Yongna, Siwei Yu, and Jianwei Ma. 2018. "Intelligent Interpolation by Monte Carlo Machine Learning." Geophysics 83 (2): V83–V97.

Kapoor, Gagan. 2017. "Estimating Pore Fluid Saturation in an Oil Sands Reservoir Using Ensemble Tree Machine Learning Algorithms." Saint Mary's University.

Khan, Mohammad Rasheed, Zeeshan Tariq, and Abdulazeez Abdulraheem. 2018. "Machine Learning Derived Correlation to DetermineWater Saturation in Complex Lithologies." Society of Petroleum Engineers.

Lewis, Winston, and Denes Vigh. 2017. "Deep Learning Prior Models from Seismic Images for Full-Waveform Inversion." In 1512–7. SEG Technical Program Expanded Abstracts 2017.

Russell, Brian, Christopher Ross, and Larry Lines. 2002. "Neural Networks and Avo." The Leading Edge 21 (3): 268–314.

Silva, Adrielle, Irineu Lima Neto, Abel Carrasquilla, Roseane Misságia, Marco Ceia, and Nathaly Archilha. 2014. "Neural Network Computing for Lithology Prediction of Carbonate- Siliciclastic Rocks Using Elastic, Mineralogical and Petrographic Properties." In 1055–8. 13th International Congress of the Brazilian Geophysical Society; EXPOGEF, Rio de Janeiro, Brazil, 26-29 August 2013.

Smith, Kenneth. 2017. "Machine Learning Assisted Velocity Autopicking." In 5686–90. SEG Technical Program Expanded Abstracts 2017.

Tittman, J., and J. S. Wahl. 1965. "The Physical Foundations of Formation Density Logging (Gamma-Gamma)." Geophysics 30 (2): 284–94.

Wrona, Thilo, Indranil Pan, Robert L. Gawthorpe, and Haakon Fossen. 2018. "Seismic Facies Analysis Using Machine Learning." Geophysics 83 (5): O83–O95.

Yue, Ming, Weiyao Zhu, Hongyan Han, Hongqing Song, Yunqian Long, and Yu Lou. 2018. "Experimental Research on Remaining Oil Distribution and Recovery Performances After Nano-Micron Polymer Particles Injection by Direct Visualization." Fuel 212: 506–14.

Zhang, Licheng, and Cheng Zhan. 2017. "Machine Learning in Rock Facies Classification: An Application of Xgboost." In 1371–4. International Geophysical Conference, Qingdao, China, 17-20 April 2017.

Zhang, Qitao, ChenjiWei, YuheWang, Shuyi Du, Yuanchun Zhou, and Hongqing Song. 2019. "Potential for Prediction of Water Saturation Distribution in Reservoirs Utilizing Machine Learning Methods." Energies 12: 3597.