geoconvention
Virtual Event
September 13 - 15 2021

Lessons Learned, Pitfalls and Feature Engineering for FORCE 2020: Log Facies Classification using Machine Learning.

David J. Emery, Marcelo Guarido, Daniel Trad and Kristopher Innanen
University of Calgary, CREWES Consortium

Summary

FORCE 2020: Machine Predicted Lithology challenge was a classification contest using well logs from the North Sea's Norwegian coast (Guarido et al. 2020). This talk focus on the lessons learned on automated well log preparation (feature engineering), effects on the various machine predictions, and the pitfalls of the provided lithofacies classes.

Overview

As part of the contest, 108 wells were provided: 10 with only the wire-line logs and 98 with lithofacies classes. The objective was to create a Machine Learning (ML) script that would determine lithofacies from well logs, with the 98 wells for a training set and the additional 10 for an intra-contest leader board. The contest's winner was selected by running the contestants' ML scripts over another set of well logs that were not provided.

For the contest, we used different ML methods from sklearn, such as Logistic Regression (to create a baseline), Naïve Bayes, Random Forest, and Gradient Boosting. All these methods require completed data, for which we used IterativeImputer using the Bayesian Ridge option after using different strategies for feature engineering (figure1).

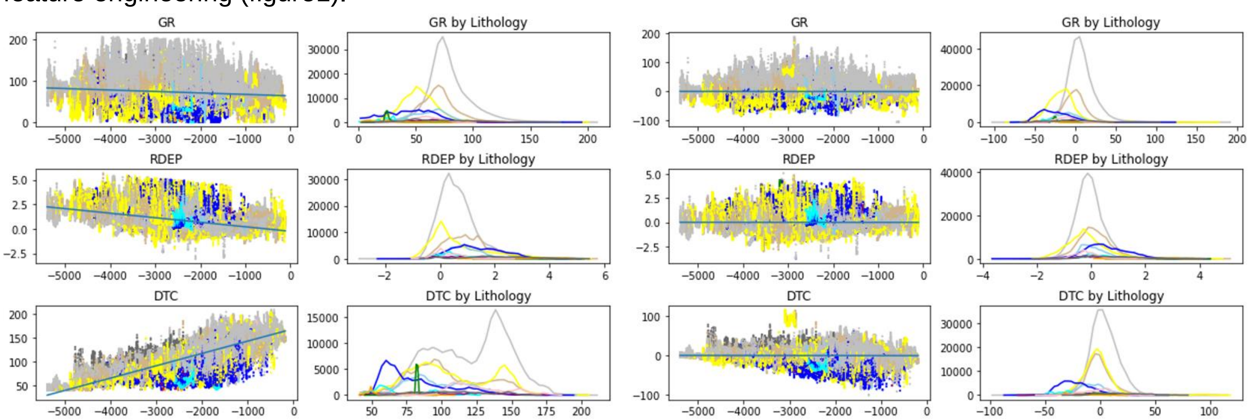

Figure1: Input and after feature engineering; display on the left is the coloured by lithology and on the right histogram for each input lithofacies. Bayesian statistics significantly improved with resulting in better separation of lithology.

Coupled with the problems of missing log data, was the Lithofacies provided was highly unbalanced, with 62% of the input samples classed as shale and 7 of the 12 classes represented less than 4% of the provided dataset. Results from ML differ on how this imbalance was dealt with; when calculated using the unbalance data overly focused on predicting shale, to when weighting the classes by the inverse of their frequency (balanced) aid estimation of the minor classes but at the trade-off of decrease the accuracy for the most frequent. We investigated the role of various balancing strategies using the Python library imblearn and sub-sampling the input lithofacies using K-mean cluster analysis.

For feature engineering, we took advantage that >85% was siliciclastic to rescaling to consistent units, remove log spikes, bulk shift for variation in water depth, and removing background depth trend.

A third problem emphasized from our investigation of class imbalance was the pitfall of mixed mineralogy within the classes, similar mineralogy shared by multiple classes (Limestone vs Chalk), and noise/outliers within each class. The contest score focused on solving this confusion between the classes.

Workflow

The baseline Logistic Regression analysis was performed using the unedited input data: first for all logs, including the X, Y & GROUP, and a second analysis using only the dominant mineralogical logs (GR, RES, DTC, NPHI, RHOB, PEF, Z).

Next, we performed ML Naïve Bayes, Random Forest, and Gradient Boosting analysis following a simple set of feature engineering compensated for differences in units, water depth, outlier removal, and normalization. As ML required completed data estimating the missing log data improved after the feature engineering. The non-linear nature of the 3 ML methods showed improvement over Logistic Regression which improved further when handling class imbalance by using either a semi-balanced or a fully balanced approach (figure 2).

As the input data covers 5000 m, a third run was conducted to investigate the role porosity variation with depth was removed from the data. Finally, uncertainty in the homogenetic nature of the input lithofacies we created subclass breaking the Shale, Sandstone, SS-SH, Marl & Limestone classes.

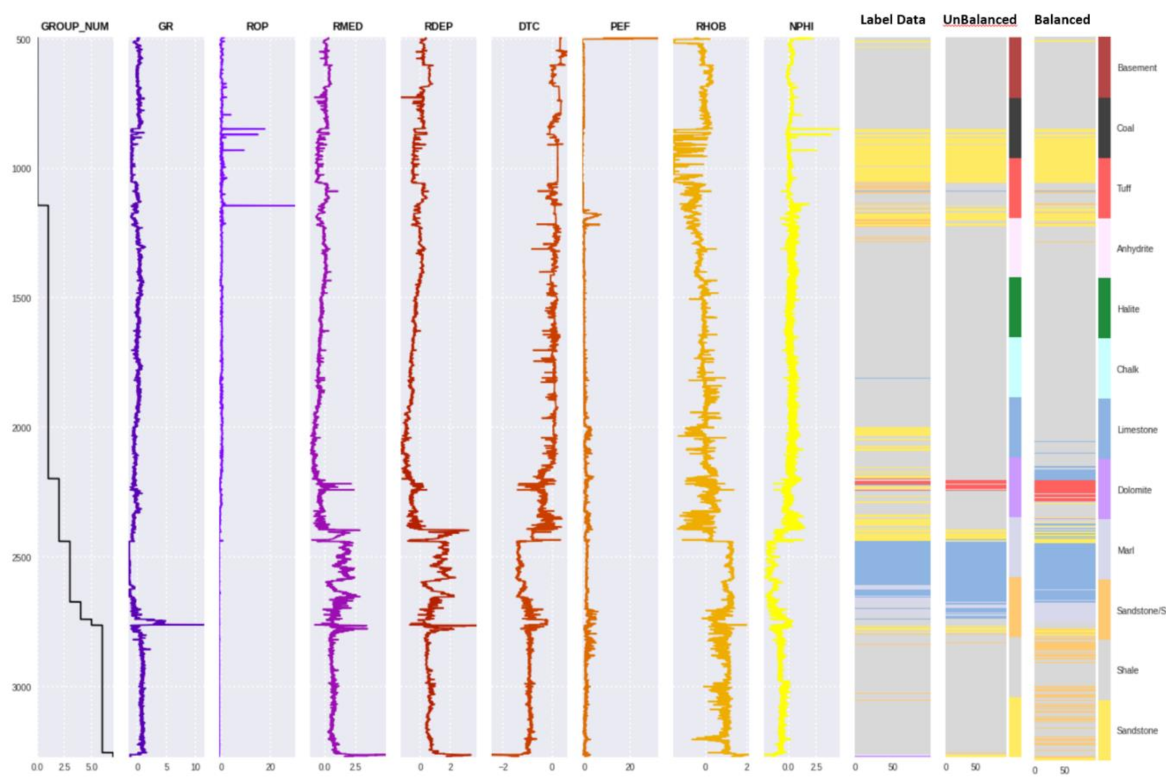

Figure2: Prediction result for unbalanced (middle) and balanced (left) using for the challenge.

Results

The Logistic Regression analysis on the raw data was extremely low (R^2 of 0.08), simple feature engineering improved the correlation and further dealing with the imbalanced nature of the data (R^2 of 0.32). The non-linear approaches of Naïve Bayes, Random Forest and Gradient Boosting increase the R^2 again by ~0.1 and the inclusion of a geological framework (X, Y, Group) was significant in both estimating the missing well data and the Lithofacies, increasing the R^2 by another factor of 0.1.

Lithofacies results for Halite, Chalk & Dolomite for all approaches were mixed to poor. The dominant class of Shale and Sandstone was highly dependent on the strategy used to handle the imbalance. While the R^2 improved from dealing with the imbalance in lithofacies, the contest score, which focused on the confusion matrix, actually became worst. Our best estimate by the end of the contest was a R^2 of 0.56 with a contest score of -0.59 (closer to 0 is better) using a stack (soft, averaging the probabilities) of the three non-linear ML approaches, and our result was mid-range of the contestants.

Since the contest, we have continued to improve feature engineering, investigating compensating for the depth trends, using a semi-balanced approach, and sub-class of the mixed mineralogy class (Shale, Sandstone-Shale, Marlstone and Limestone). These approaches have improved the estimation of missing data, the contest score, and the R^2.

Conclusion

While stacking the 3 ML methods has given the best results, the Gradient Boosting results appear the best of the 4 methods used and will be the concentration of the talk. Three pitfalls became evident: the feature engineering required to estimate the missing data, the mixed mineralogy in the provide Lithofacies, and metrics for scoring success.

Finding a good machine learning solution to Petrophysical analysis is highly probable, but finding the right approach for data preparation, mineralogy determination, and finally, facies classification will take time. Our present work would indicate that having a geological framework should significantly aid feature engineering. A single-point estimator may be limited to determining a class with similar or mixed mineralogy.

Acknowledgements

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13 and CRDPJ 543578-19, and the financial support from Canada First Research Excellence Fund.

References

Guarido, M., Emery, D. J., Macquet, M., Trad, D. O., and Innanen, K. A. H., 2020, The Pitfalls and Insights of Log Facies Classification for a Machine Learning Contest: CREWES Research Report, 32, 18.