Machine Learning Strategies to Perform Facies Classification

Marcelo Guarido
CREWES (University of Calgary) and Machine Learning Team at Verdazo Analytics

Summary

Lithological facies classification is the process to determine rock lithology by analyzing indirect measurements such as well logs. Usually classification is done manually by an experienced geological interpreter. In this work, we present an method for automated facies classification using feature engineering and ensemble classifiers (machine learning). Facies logs from several interpreted wells are used to train multiple multiclass machine learning models. Predictions from these models using both raw and processed log features are then combined and compared with facies labels in a blind well. By engineering new input features from the original well logs using methods such as local gradient computation, coordinate transformations, and clustering analysis, and by combining the results from multiple models, prediction accuracy from the ensemble classifiers increases from 47% to 64%.

Introduction

Machine learning is a computer science field with growing attention on the last ten years, thanks to the increase of computational power and readily-available open source libraries in recent years. Fed by a reasonable amount of data, machine learning algorithms can make predictions or recognize patterns. The goal of this work is to train a machine learning algorithm to predict rock types by feeding the algorithm with well logs (input data) and core-derived facies interpretations (labels), so that rock types can be predicted from wireline logs alone. Machine learning algorithms for facies classification have been performed in the past with different approaches, such as the use of ensemble classifiers (Bestagini et al., 2017; Zhang and Zhan, 2017; Caté et al., 2017, Guarido, 2018), neural networks (Silva et al., 2014), and support-vector machines (Caté et al., 2017; Alexsandro et al., 2017; Wrona et al., 2018).

Guarido (2018) shows that, in the absence of feature engineering, individual machine learning algorithms for classification can lead to adequate predictive performance. However, more powerful predictive models can be created by forming ensembles of different classifiers.

Method and Data

Hall (2016) proposed a facies classification contest using machine learning, with the results printed in Hall and Hall (2017). Data published online for open use were used in this work. The dataset contains ten wells, each having seven features (five well logs and two indicators) covering the zone of interest, as listed below:

1. Gamma Ray (GR)
2. Resistivity (ILD)

3. Photoeletric effect (PE)
4. Neutron-density porosity difference (DeltaPHI)
5. Average neutron-density porosity (PHIND)
6. Nonmarine/marine indicator (NM_M)
7. Relative position (RELPOS)

Different rock types, identified on core, are labelled in the dataset and are represented (classified) by an integer value from 1 to 9, as shown in Table 1. Note the column adjacent facies.

Table 1: Facies labels and their descriptions. Note the ***Adjacent Facies*** column, which lists facies that are found adjacent to that labelled in the ***Facies*** column

Facies	Description	Label	Adjacent Facies
1	Nonmarine Sandstone	SS	2
2	Nonmarine coarse siltstone	CSiS	1,3
3	Nonmarine fine siltstone	FSiS	2
4	Marine siltstone and shale	SiSh	5
5	Mudstone	MS	4,6
6	Wackestone	WS	5,7,8
7	Dolomite	D	6,8
8	Packstone-grainstone	PS	6,7,9
9	Phylloid-algal bafflestone	BS	7,8

Guarido (2018) shows that using only the seven given features without any further treatment, the accuracy of the predicted facies in a blind well is 47% using a gradient boosting classifier. However, after estimating missing data using linear regression and by creating new features from the original features (computing polar coordinates, local log gradients, and clustering analysis), accuracy improves to 60% on the same blind well.

This work follows a similar approach to Guarido (2018) for feature augmentation, but the missing data (two wells missing the PE log) was predicted using a *gradient boosting regressor*. Furthermore, instead of using gradient boosting as the lone classifier algorithm, many other classifiers, such as logistic regression, support vector machine (SVM), random forest, gradient boosting, and neural networks, are tuned and then combined to produce a more powerful ensemble classifier.

Results

Of the ten wells initially provided, one provided incomplete, unordered data and brought no reliable information. This well was removed from the dataset. Of the nine remaining wells, two had missing PE logs. To fill this missing data, a gradient boosting regressor was trained to predict the PE log from the GR, ILD, DeltaPHI, and PHIND logs using six of the remaining complete wells. The PE log from the seventh complete well used to validate the PE log predictions and select the best parametrization of the gradient boosting regressor. The trained regressor was used to predict the PE log on the two wells missing valid

PE information. Figure 1 shows the predicted PE in the validation well (left) and on the well "Alexander D" (right), one with the missing PE.

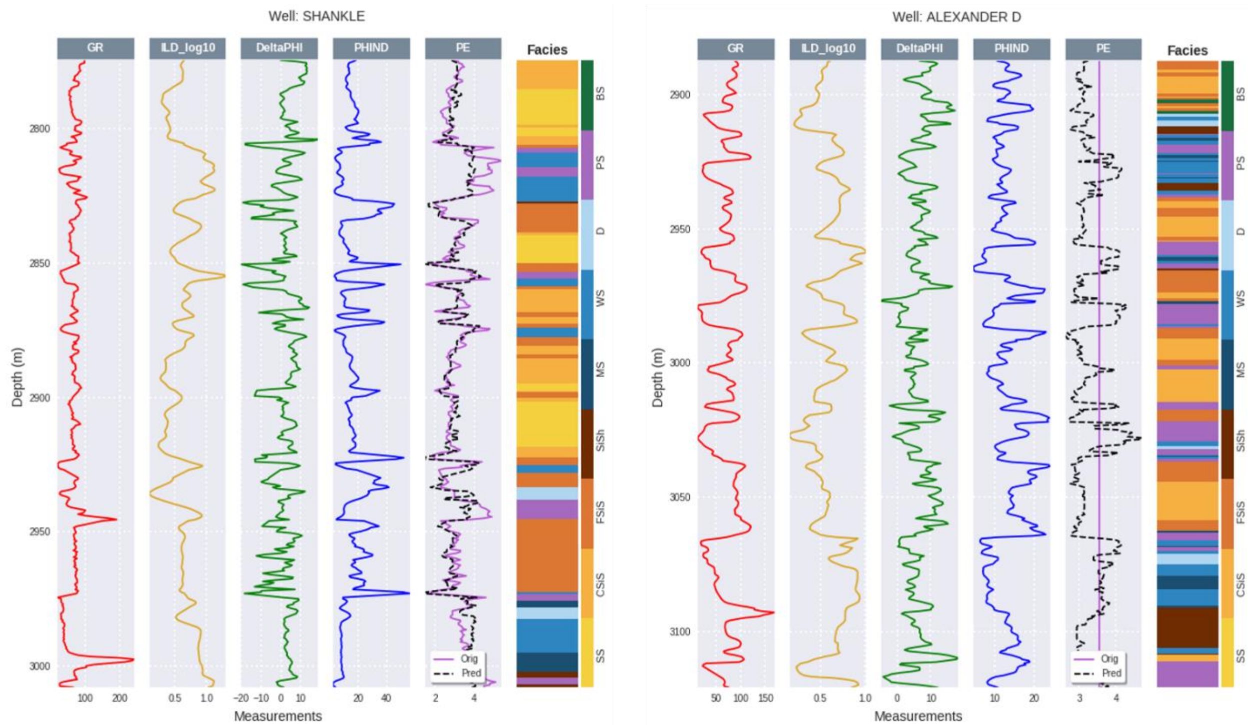
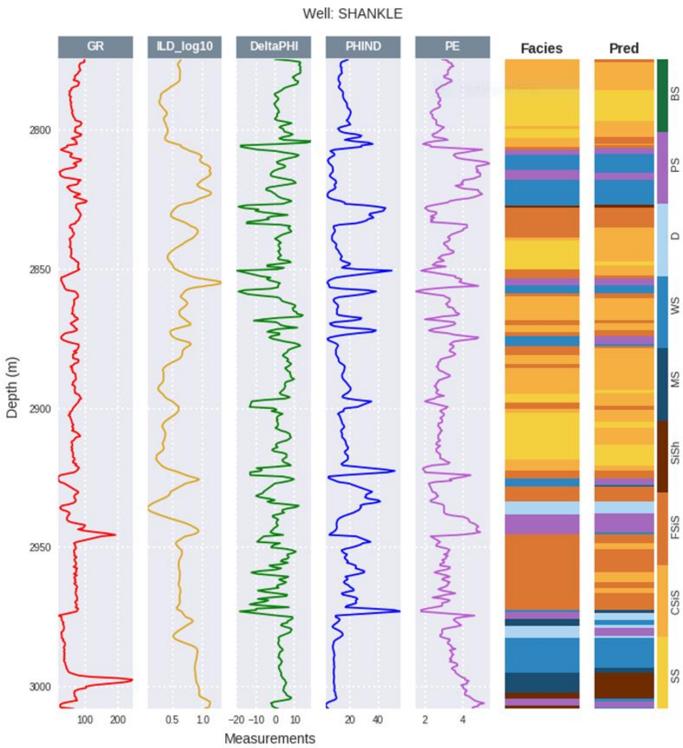


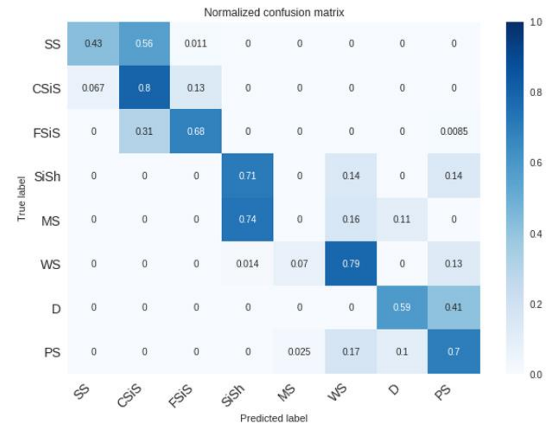
Figure 1: Example of the gradient boosting regression algorithm to complete missing data on validation well (left) and on one of the wells with missing PE (right).

After augmenting the input data to fill in missing values, a typical subsequent step would be to train a model to predict the facies classes. However, Guarido (2018) shows that create new features from the existing ones improves the predictions. As Guarido (2018) shows, circular patterns exist on certain property cross plots. Polar coordinates of those property pairs are computed to convert circular patterns in Cartesian space to linear patterns in polar-coordinate space, which are easier for a classifier to separate. A smooth gradient was also computed for each of the input logs, to identify local vertical trends. In addition, k-means classification was used to generate groups of similar input log values, and these group identifiers were fed as engineered features to the predictive models.

Different classifiers for facies prediction were tested and tuned. Figure 2 contains a table with the accuracies of the tested models. Tree-based methods (random forest and gradient boosting) attained the top accuracies. However, when checking balanced accuracies, logistic regression is the best model, followed closely by gradient boosting. The interpretation is that tree-based models were very effective in predicting highly-sampled facies, while logistic regression predicted all rock types in a more balanced matter. SVM, neural networks, and random forest were ineffective in predicting facies represented by only a few samples in the training set.

The best predictive power is achieved by combining multiple classifiers. Figure 2 compares the predicted vs. actual facies, and the associated confusion matrix, for predictions on the blind well, which was not used





Model	Accuracy Score (%)	Balanced Accuracy Score (%)
Logistic Regression	58.4	57.1
SVM	56.8	41.1
Neural Networks	58.1	48.3
Random Forest	60.0	47.7
Gradient Boosting	62.1	56.0
Ensemble Classifiers	64.1	58.7

Figure 2: Result with the ensemble classifiers. The confusion matrix shows the prediction quality for each class, where in the horizontal direction shows the predicted labels, and the vertical is the true rock type. The table lists each model accuracy (just the proportion of correct predictions) and balanced accuracy (accuracy normalized by class).

in training the model. Most facies are correctly predicted. Sandstone (SS) was mainly misclassified as coarse siltstone due to their similarity in the logs. Mudstone had no correct predictions, as it is the least sampled: the model has insufficient data from which to learn a reliable pattern to identify rock type.

Conclusions

A classification strategy is presented to improve facies predictions using well logs. Engineering new features to help a machine learning model identify patterns, combined with an ensemble of tuned classifiers, proved to be a successful pipeline to predict rock types. Less reliable predictions were obtained for rock types under-represented in the training dataset.

Acknowledgements

I thank Verdazo Analytics, CREWES, NSERC and CFREF for the support and financing of the research. I also thank the Machine Learning team at Verdazo Analytics (Anton Biryukov, Brian Emmerson, Manoochehr Akhlaghinia, and Tyler Schlosser), Erick Gomes Anatácio, and Soane Mota dos Santos for suggestions, tips, and productive discussions.

References

Alexsandro, G. C., da P. Carlos, A. C., and Geraldo, G. N., 2017, Facies classification in well logs of the Namorado oilfield using Support Vector Machine algorithm, 15th International Congress of the Brazilian Geophysical Society; EXPOGEF, Rio de Janeiro, Brazil, 31 July-3 August 2017, 1853–1858.

Bestagini, P., Lipari, V., and Tubaro, S., 2017, A machine learning approach to facies classification using well logs, SEG Technical Program Expanded Abstracts 2017, 2137–2142.

Caté, A., Perozzi, L., Gloaguen, E., and Blouin, M., 2017, Machine learning as a tool for geologists: The Leading Edge, 36, No. 3, 215–219.

Guarido, M., 2018, Facies classification with features engineering, clustering, and gradient boosting trees: CREWES Research Report, 30, 13.1 – 13.24.

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, 35, No. 10, 906–909.

Hall, M., and Hall, B., 2017, Distributed collaborative prediction: Results of the machine learning contest: The Leading Edge, 36, No. 3, 267–269.

Silva, A., Neto, I. L., Carrasquilla, A., Misságia, R., Ceia, M., and Archilha, N., 2014, Neural network computing for lithology prediction of carbonate- siliciclastic rocks using elastic, mineralogical and petrographic properties, 13th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 26–29 August 2013, 1055–1058.

Zhang, L., and Zhan, C., 2017, Machine Learning in Rock Facies Classification: An Application of XGBoost, International Geophysical Conference, Qingdao, China, 17-20 April 2017, 1371–1374.