Facies Characterization of a Reservoir in the North Sea Using Machine Learning Techniques

Peipei Li (peipei@stanford.edu); Yuran Zhang (yuranz4@stanford.edu)

Introduction

In oil and gas reservoirs, the subsurface is largely heterogeneous and oil is not present everywhere. Although well logging techniques allow us to directly differentiate different facies at the wellbore via gamma ray, it is not able to characterize facies in the reservoir that is further away from the wellbore .

We propose using machine learning techniques to perform reservoir facies classification (oil sand / brine sand / shale) based on non-gamma-ray well log data (V_p, V_s , density, etc.). The result can be applied to a 3D reservoir with seismic reflection data.

Data Processing

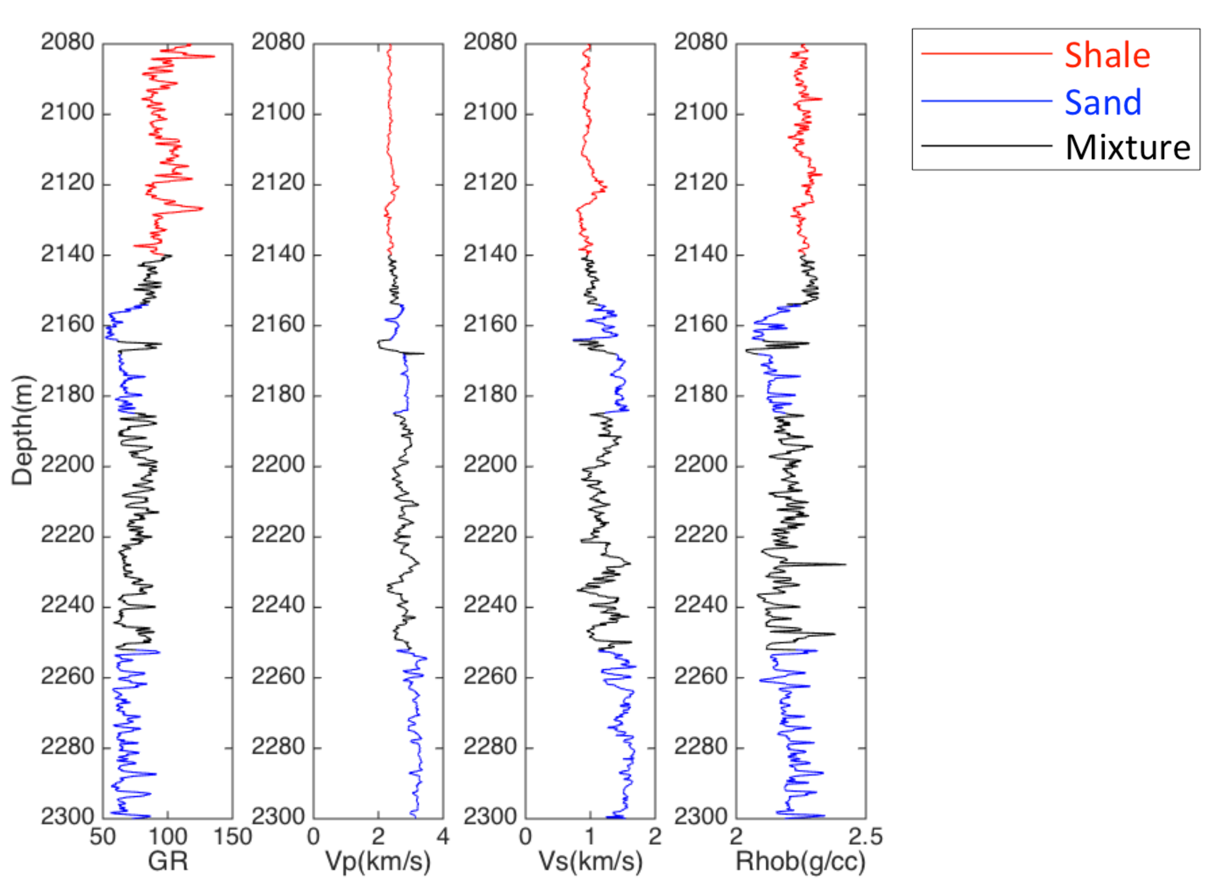
Data Set:

Well log data and seismic data from a reservoir in the North Sea.

Facies Identification:

Two facies (shale and brine sand) were identified using GR log; A third facies (oil sand) was created using fluid substitution.

Total number of samples: 1377.



Feature Creation

In addition to the three features shown above (V_p, V_s , density), six more features were created using rock physics function. In total, nine features were used in modeling: P-wave velocity, S-wave velocity, density, shear modulus, bulk modulus, P-wave impedance, S-wave impedance, Poisson's ratio and Lame's coefficient.

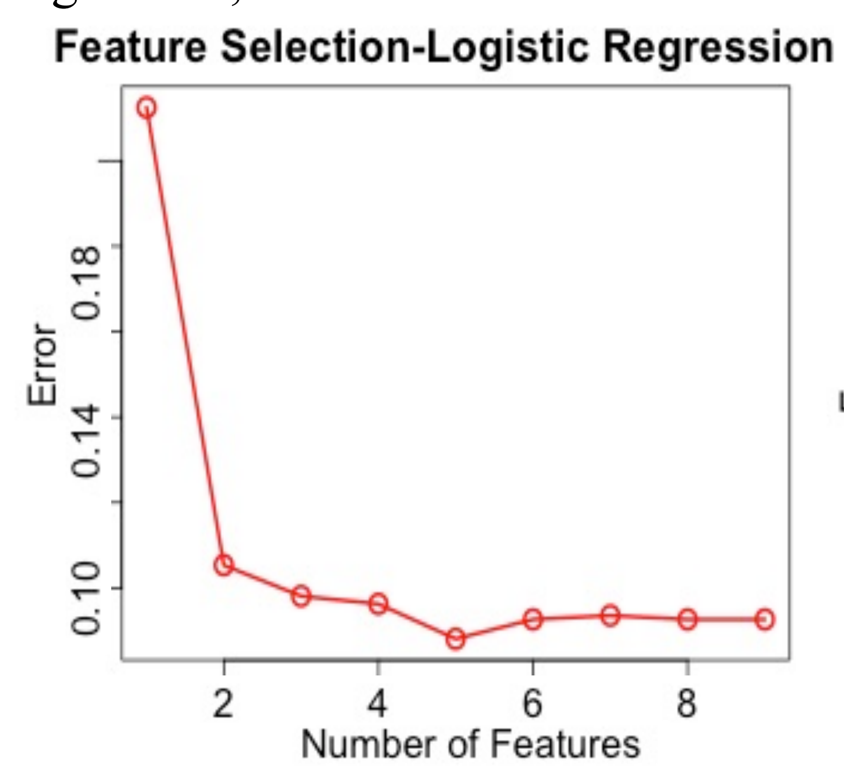
Implementation of ML Algorithms

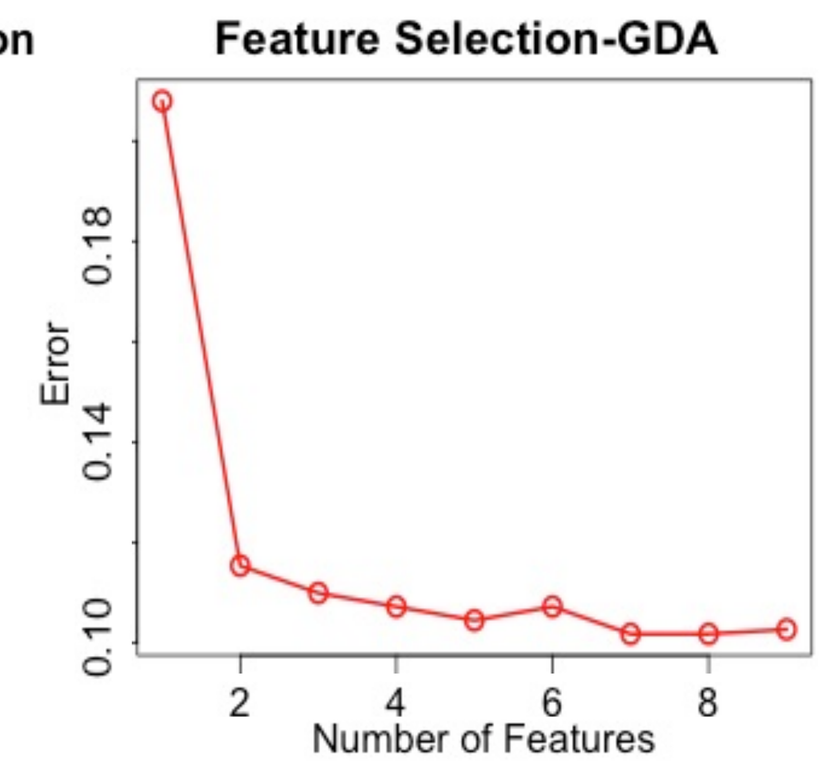
Two linear classification techniques were first implemented (in R):

Softmax Regression ("multinom" from "nnet" library) is a generalization of logistic regression applicable to cases with more than two labels.

Gaussian Discriminant Analysis (GDA) ("lda" from "mass" library) is a generalized linear model, which assumes normal distribution of the data.

For both methods, forward search feature selection was performed, where each feature combination in the forward search was evaluated with k-fold cross validation (k=10). As shown in the error vs #features plot below, the testing error is lowest when 5 features were evaluated for softmax regression, and 7 features for GDA.

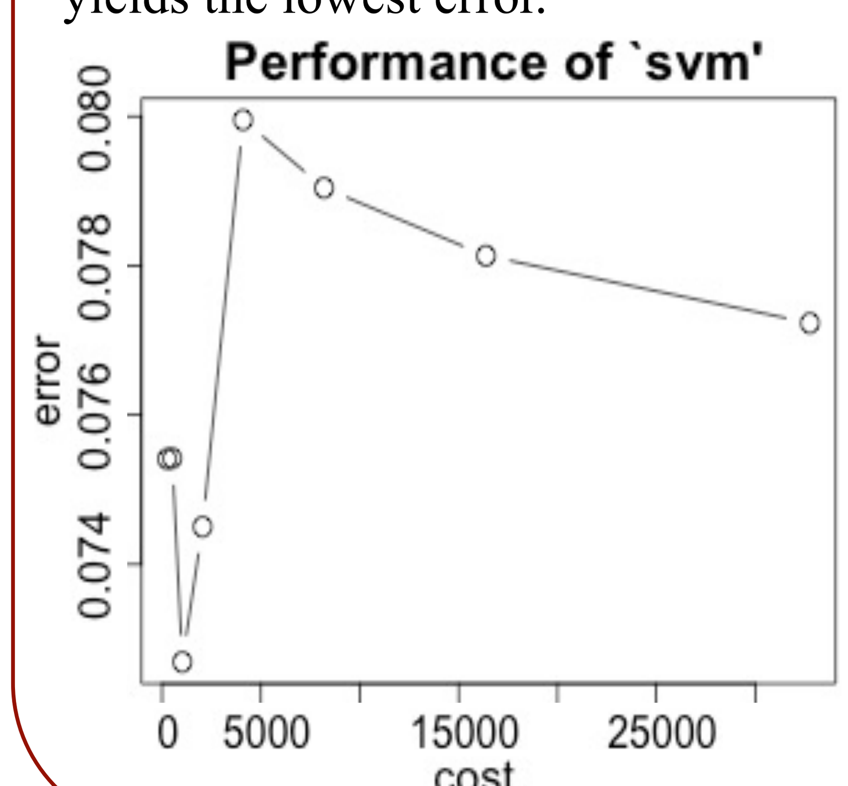


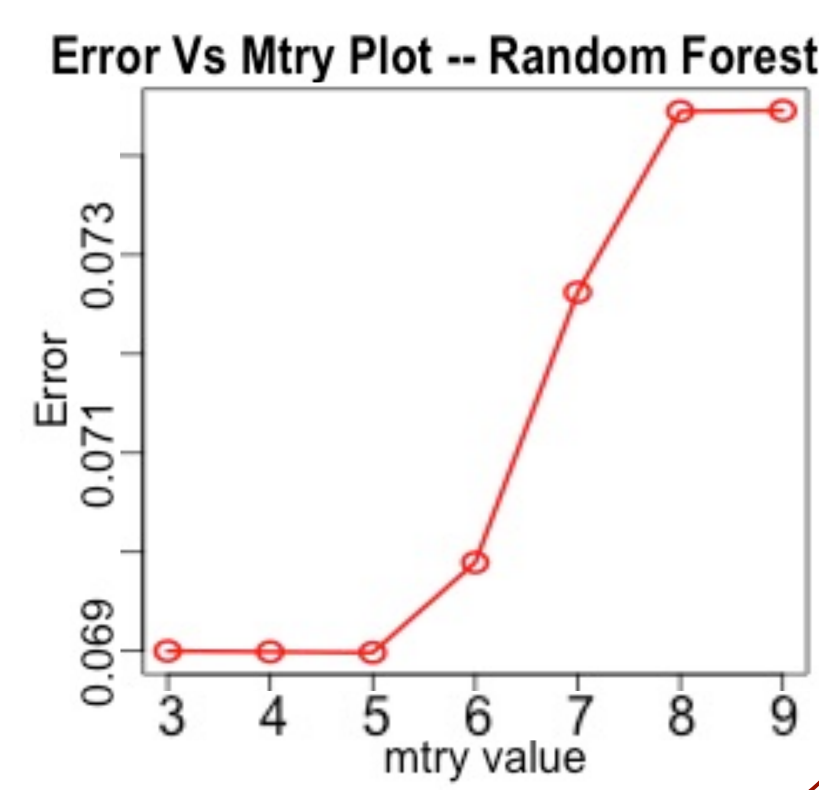


Two nonlinear classification techniques were also implemented (in R):

Support Vector Machine (SVM) ("svm" from "e1071" library) is a maximum margin classifier. The kernel function used in this project is Gaussian radial basis function. A range of the cost parameter C is evaluated, and the result is optimal when C = 1024.

Random Forest ("randomForest" from "randomForest" library) is a special case of bagging methods for decision trees. We used k-fold cross validation (k=10) to determine the best "mtry" value for classification, where "mtry" denotes the #variables randomly sampled as candidates at each split. As indicated in the figure below, mtry = 5 yields the lowest error.





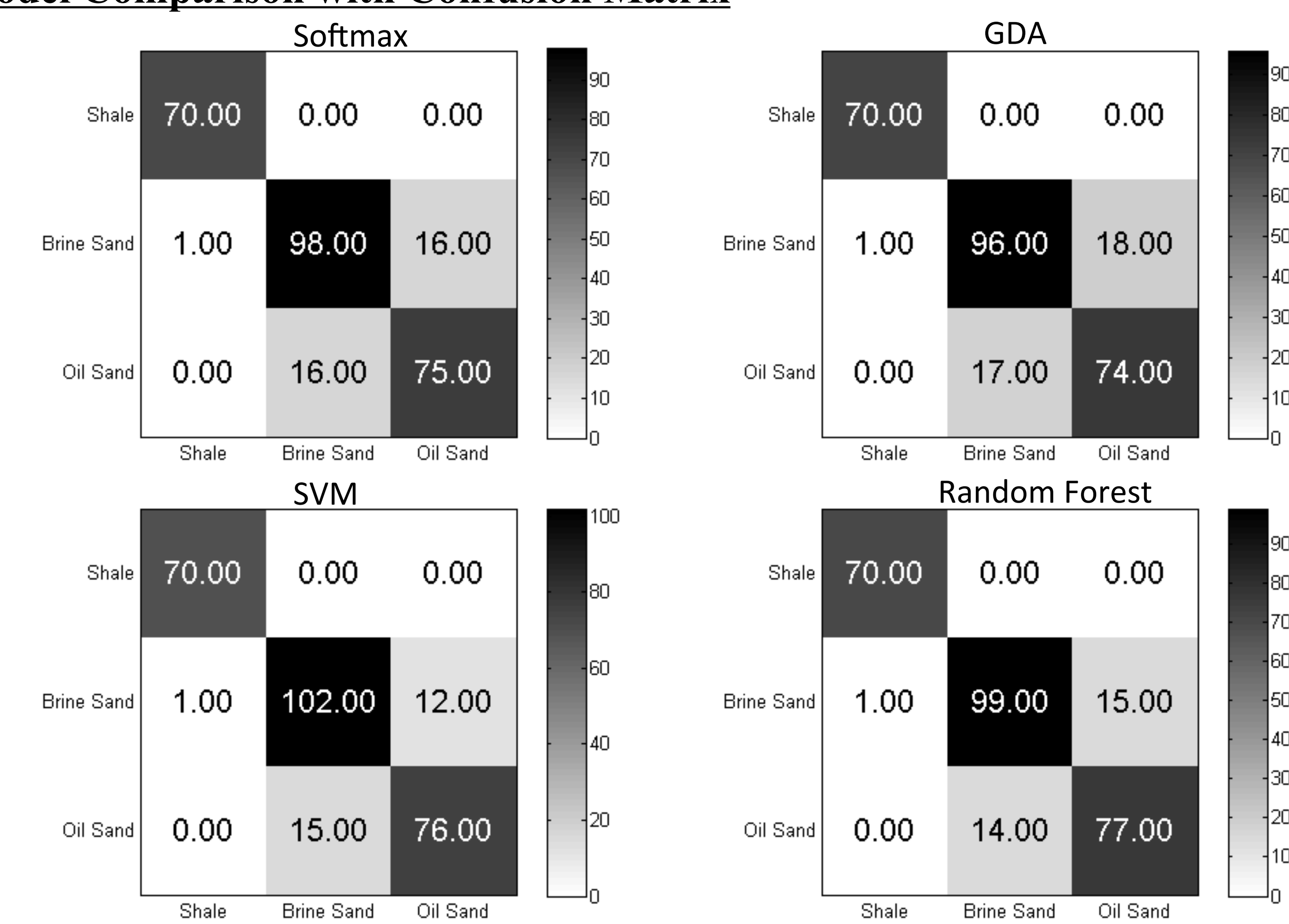
Results & Discussion

Model Comparison with MSE

Methods	Training Error	Testing Error
Softmax Regression	0.086	0.12
GDA	0.107	0.13
SVM	0.051	0.101
Random Forest	0.065	0.109

- Each model was run with selected parameters.
- Non-linear models are generally better than linear models.
- Testing errors are slightly higher than training errors, indicating proper model complexity.

Model Comparison with Confusion Matrix



- Results are consistent with model comparison using MSE.
- All models can at least discriminate sand from shale.
- SVM achieves best classification results, especially in terms of false positives of oil sand.

Future Work

With the conclusion above, the next step is to apply the SVM to 3D dataset for facies classification.

However, we only have P-wave impedance and S-wave impedance inverted from seismic data available for use. As a reference, we built a softmax regression model using these two features only and it showed an error of 0.23. So we think getting more seismic attributes could improve the classification significantly.

Reference:
Avseth, P., et al. "Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system." Geophysics 66.4 (2001): 1157-1176.