SEISMIC FACIES CLASSIFICATION USING A RANDOM FOREST ALGORITHM

Yuji Kim*[1], Robert Hardisty[1], Emilio Torres[1], Kurt J. Marfurt[1]
[1]University of Oklahoma, OK, USA

Abstract

In this study, we use a random forest learning algorithm to predict seismic lithofacies aided by wireline logs and stratigraphical interpretation. The random forest algorithm is a tree-based classifier that is an alternative to neural-network and support vector machine based algorithms. The first goal of this study is to evaluate the ability of a random forest learning algorithm in classifying seismic facies, The second goal is to determine the sensitivity of this classification to each of the input attributes. This sensitivity analysis will not only help choose the most attribute attributes but will also reduce the amount of computations needed. We are able to differentiate limestone and shale facies in a Fort Worth Basin survey and salt, MTC, and conformal sediment facies in a Gulf of Mexico survey.

Introduction

Automated seismic facies classification is gradually becoming more common as part of workflows in the E&P industry, and has seen improved success since early work by Coléou. et al. (2003). Several algorithms have been developed to automate seismic facies classification, most of which are based on some form of machine learning. Machine learning can be either unsupervised or supervised. Unsupervised learning, e.g. *k*-means clustering, principal component analysis (PCA), and self-organizing maps (SOM) require little interpreter input other than the selection of appropriate attributes. Although all major facies in the volume are classified, the algorithm gives no indication of the identity of any given facies. In contrast, supervised classification such as multilevel feed forward neural networks, support vector machines, and random forest decision trees assign each voxel to facies defined by the interpreter using a "training" subset of the data. Well log and core data provide not only the training but also validation data used to quality control the prediction. We apply a random forest algorithm which is an ensemble of deicision trees, trained via a "bagged" (or "boosted") method. The random forest or ensemble methods indicate which trees are considered "strong" learners and "weak" learners prior to combining them.

Theory

Single decision tree and random forests

Classification and regression tree (CART) is a machine leaning technique. Breiman et al.'s (1984) CART algorithm splits each internal node in the tree using the "Gini" impurity $i(\tau)$ given by

$$Gini\ i(\tau) = 1 - \sum_{\theta=1}^{k} p\,(\theta|\tau)^2$$

where k is the number of classes, and $p(\theta \mid \tau)$ is the probability of class θ at node t .
Prediction for *N* trees can be made by averaging the prediction of individual trees:

$$i_{N_T}(\tau) = \frac{1}{N_T}\sum_{T} i(\tau)$$

Feature importance

Selecting appropriate features is important in machine learning algorithms. Some features are more important to classification, while others may be redundant. Reduction of dimension based on feature selection can speed up the learning process, as well as improve prediction accuracy. To evaluate feature importance, Breiman et al. (2001, 2002) defined Gini importance based on the Gini impurity as

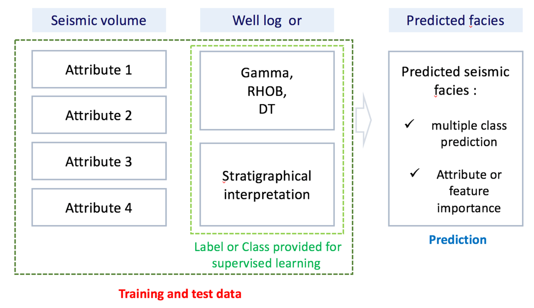
$$Gini\ Importance\ i_G(\theta) = \textstyle\sum_T \sum_\tau \Delta i_\theta(\tau, \mathrm{T}),$$

where, $\Delta i(\tau)$ is the node purity gain define as

$$\Delta i(\tau) = i(\tau) - p_l i(\tau_l) - p_r i(\tau_r).$$

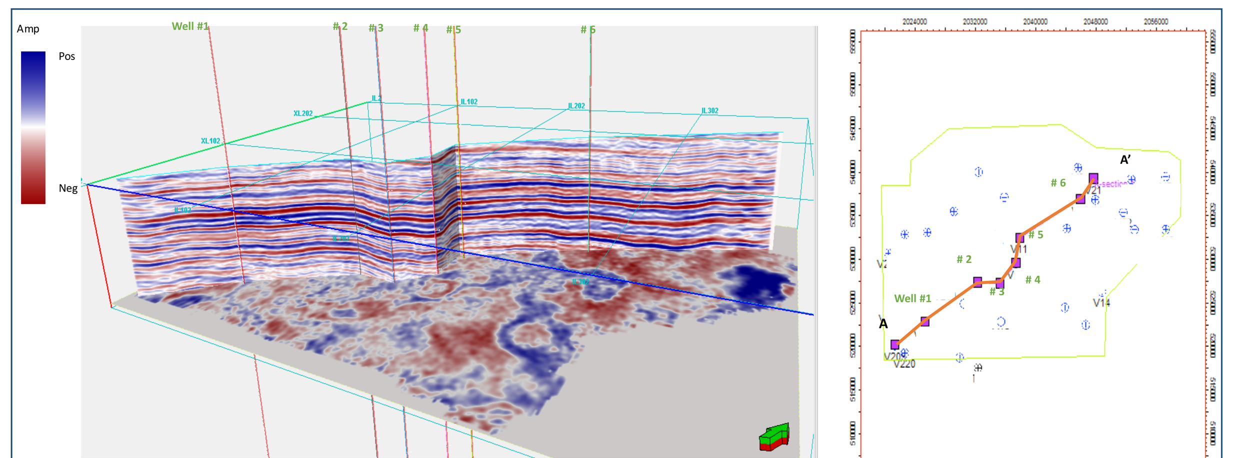
Case study 1: Fort Worth Basin Survey

- **Aim of the study:**
  - Descriminate seismic carbonate and shale facies aided by well log data training

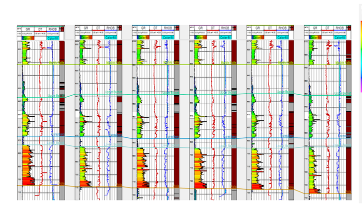


- **Training / validation set** -
  - Input: data points in 3D volumes adjacent to well log (6 inverted seismic physical properties)
  - Output : seismic facies interpreted from 6 well logs

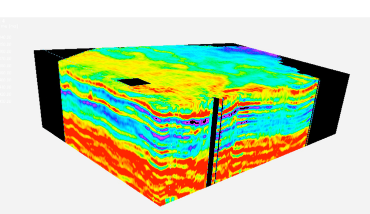
- **Training input data**



(Left) Vertical composite line AA' connecting multiple wells and a representative time slice through the seismic amplitude volume. (Right) Map of seismic survey and location of wells used intraining.



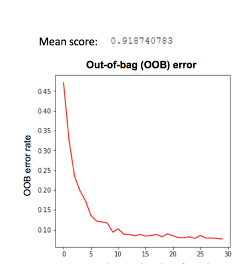
Lithofaces interpreted from six well logs



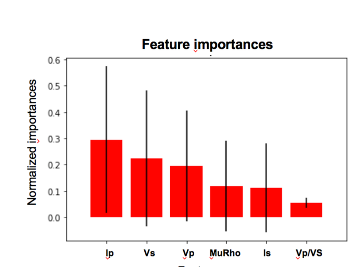
Input attribute example: P-wave impedance (Z_p)

Statistical analysis of Input training data

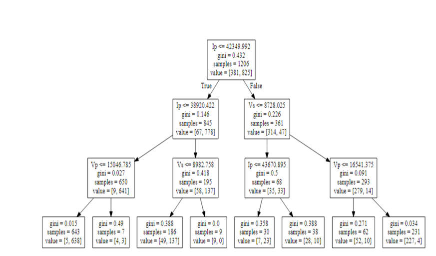
- 5 fold cross-validation with training data
  - Random forest model
  - Average accuracy: 0.9187



Number of trees versus error rate



Feature importance

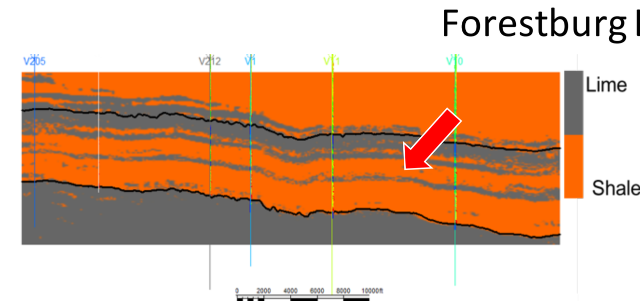


A tree diagram generated from a single tree model

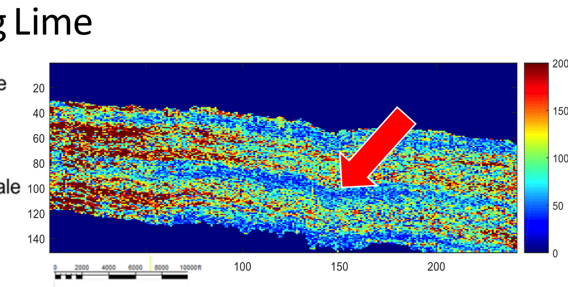
- **Predicted seismic facies in comparison to unsupervised learning algorithm (Nearest neighbors)**

- Nearest neighbors is non-parametric, unsupervised classification by finding the most similar data points in the training data

Forestburg Lime



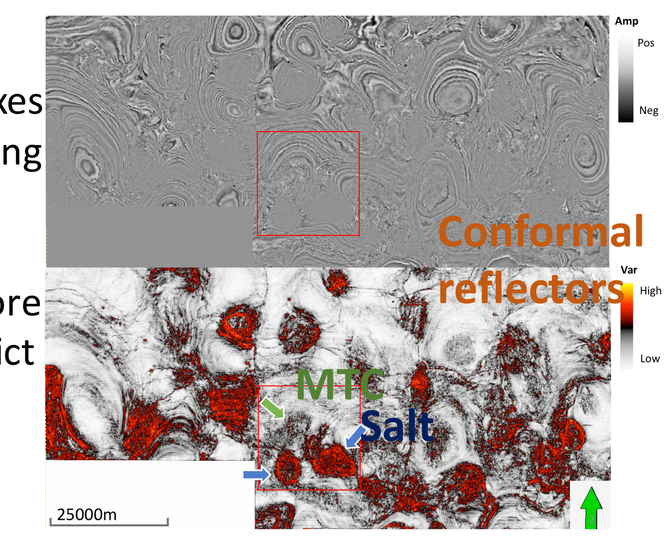
A representative vertical slice through predicted using random forest



Nearest neighbors clustering (unsupervised)

Case study 2: Gulf of Mexico Survey

- **Aim of the study:**
  - Predict seismic lithofacies (Salt diapirs, Mass transport complexes (MTC), conformal reflectors) using multiple seismic attributes.
  - Which seismic attribute are more important and effective in predict these three facies ?



Time slice through seismic amplitude (top) and variance attribute (bottom)

- **Dataset:**
  20 Seismic attributes (amplitude, geometry, texture, ..) generated from 3D seismic acquired in the Gulf of Mexico survey

- **Training /validation set**
  Salt, MTC and conformal reflectors are interpreted and cropped using polygon
  → 30,000 data points are randomly sampled
  - **Input**: 30,000 data points in 3D volumes (20 seismic attributes in five categories)
  - **Output** : seismic facies from stratigraphical interpretation

- Seismic characteristics of the target lithofacies

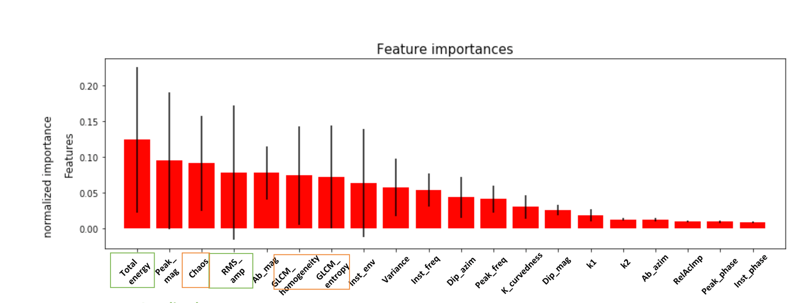
Seismic amplitude pattern and attribute anomalies of each environment		
Conformal reflectors	**Salt**	**Mass transport complexes**
Continuous High coherence Moderate values of GLCM entropy	High variance Low amplitude	Chaotic Discontinuous Hummocky Low coherence and high energy Irregular bed thickness

(Modified from Roy et al. (2013)

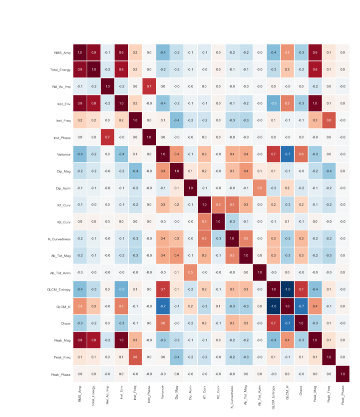
- 20 seismic attributes in five categories

Categories of seismic attributes applied to facies classification				
Amplitude attributes	**Instantaneous attributes**	**Geometric attributes**	**Texture attributes**	**Spectral attributes**
RMS amplitude Total energy Relative acoustic impedance	Instantaneous envelope Instantaneous frequency Instantaneous phase	Variance Dip magnitude Dip azimuth Most-positive curvature Most-negative curvature Aberrancy magnitude Aberrancy azimuth	Chaos GLCM entropy GLCM homogeneity	Peak magnitude Peak frequency Peak phase

- 5 fold cross-validation with 20 attributes yields accuracy : 0.92789



Feature importance based on Gini impurity decrease. As expected, phase and azimuth have little to do with facies.



Correlation between attributes

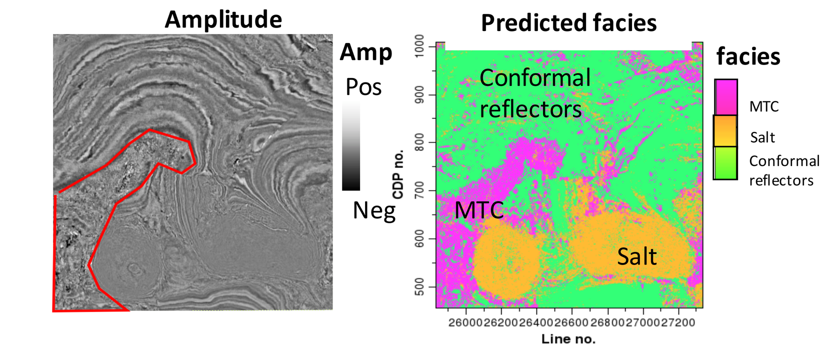
- Prediction of lithofacies with ten attributes selected

Attribute selection using prediction model (Wrapper method)

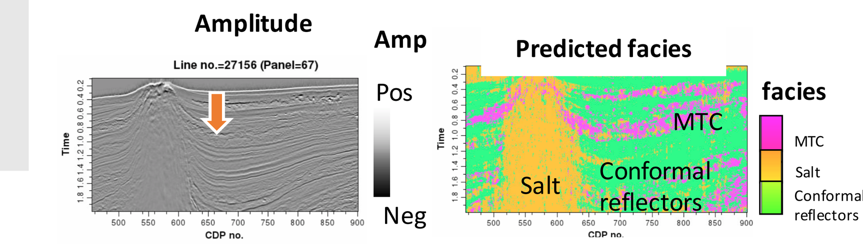
Forward selection
Prediction model: RF

- Total energy
- Chaos
- Abberancy magnitude
- Instantaneous frequency
- Dip azimuth
- Variance
- Structural curvedness
- peak frequency
- Abberancy azimuth
- Most-positive curvature

→ **Ten attribute results avg. accuracy of 0.931324**



Representative time slices through amplitude and predicted facies



Representative vertical slices through amplitude and predicted facies

- A comparison of computation time
  (RF model with n_estmators = 100, depth = 50)

Number of attributes	Number of processes	Time elapsed for training (s)	Time elapsed for prediction of 3D volume (s) (780 Mbytes)
20	1	11.15 s	$1.67E^{-5}$ / samples x (475 x 551 x 751) samples = 3276 s
10	1	8.53 s	$1.51E^{-5}$ / samples x (475 x 551 x 751) samples = 2967 s

(Intel Core i7 2.3 GHz CPU)

Conclusions

- In the Fort Worth Basin survey, two, thin layers are well defined in predicted lithofacies which correlated with a lime layer in the Upper Barnett Shale and the Forestburg Lime that separates the Upper and Lower Barnett Shale seen in the well log.
- In the Gulf of Mexico survey, amplitude and texture attributes are powerful attributes that are able to differentiate salt, MTC and conformal reflector facies.
- Misclassification occurs around fault areas where variance are high as that of MTC.
- For a given quality of classification, appropriate attribute selection can significantly reduce the computation time.

References

Coléou, T., M., Poupon and K. Azbel, 2003, Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, **22**, 942-953.

Roden, R., T., Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **3**, SAE59-SAE83.

Breiman, L., J., Friedman, C. J. Stone, and R. A. Olshen, 1984, Classification and regression trees: CRC press.

Acknowledgements

The authors thank the sponsors of OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support.

For communication email: yuji.kim@ou.edu