Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale

Yuji Kim[1], Robert Hardisty[1], and Kurt J. Marfurt[1]

Abstract

Automated seismic facies classification using machine-learning algorithms is becoming more common in the geophysics industry. Seismic attributes are frequently used as input because they may express geologic patterns or depositional environments better than the original seismic amplitude. Selecting appropriate attributes becomes a crucial part of the seismic facies classification analysis. For unsupervised learning, principal component analysis can reduce the dimensions of the data while maintaining the highest variance possible. For supervised learning, the best attribute subset can be built by selecting input attributes that are relevant to the output class and avoiding using redundant attributes that are similar to each other. Multiple attributes are tested to classify salt diapirs, mass transport deposits (MTDs), and the conformal reflector "background" for a 3D seismic marine survey acquired on the northern Gulf of Mexico shelf. We have analyzed attribute-to-attribute correlation and the correlation between the input attributes to the output classes to understand which attributes are relevant and which attributes are redundant. We found that amplitude and texture attribute families are able to differentiate salt, MTDs, and conformal reflectors. Our attribute selection workflow is also applied to the Barnett Shale play to differentiate limestone and shale facies. Multivariate analysis using filter, wrapper, and embedded algorithms was used to rank attributes by importance, so then the best attribute subset for classification is chosen. We find that attribute selection algorithms for supervised learning not only reduce computational cost but also enhance the performance of the classification.

Introduction

In the exploration and production industry, automated seismic facies classification is gradually being integrated into common workflows. Several machine-learning algorithms, such as self-organizing maps (SOMs) and K-means clustering, have been applied to automate seismic facies classification, and they are available in several commercial interpretation software packages. A great number of different seismic attributes can be used as input to machine-learning algorithms for classification and pattern recognition. However, some attributes express geologic or depositional patterns more effectively than others. For instance, the envelope (reflection strength) is sensitive to changes in acoustic impedance and has long been correlated to changes in lithology and porosity (Chopra and Marfurt, 2005). In many cases, the instantaneous frequency enhances interpretation of vertical and lateral variations of layer thickness (Chopra and Marfurt, 2005). Coherence measures lateral changes in the seismic waveform, which in turn can be correlated to lateral changes in structure and stratigraphy (Marfurt et al., 1998). Exploration generates large amounts of seismic data, and many attributes generated may be highly redundant. Adding to this problem, the original seismic amplitude data (and therefore the subsequently derived attributes) may contain significant noise (Coléou et al., 2003). Therefore, understanding the nature of seismic attributes is of crucial importance for providing the most reliable classifications.

According to the Hughes phenomenon, adding attributes beyond a threshold value causes a classifier's performance to degrade (Hughes, 1968). Several studies found that dimensionality reduction in machine-learning problems reduces computation time and storage space as well as having meaningful results for facies classification (Coléou et al., 2003; Roy et al., 2010; Roden et al., 2015). Principal component analysis (PCA) is one of the most popular methods, reducing a large multidimensional (multiattribute) data set into a lower dimensional data set spanned by composite (linear combinations of the original) attributes, while preserving variation. SOM also creates a lower dimensional representation of high-dimensional data to aid interpretation. PCA and SOM are types of unsupervised learning, in which the goal is to discover the underlying structure of the input data.

[1]The University of Oklahoma, ConocoPhillips School of Geology and Geophysics, Norman, Oklahoma, USA. E-mail: yuji.kim@ou.edu (corresponding author); bob@ou.edu; kmarfurt@ou.edu.

Manuscript received by the Editor 17 December 2018; revised manuscript received 14 March 2019; published ahead of production 29 May 2019; published online 23 August 2019. This paper appears in *Interpretation*, Vol. 7, No. 3 (August 2019); p. SE281–SE297, 16 FIGS., 9 TABLES.
http://dx.doi.org/10.1190/INT-2018-0246.1.

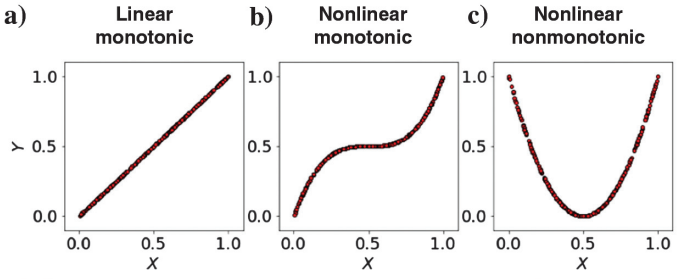
Roden et al. (2015) use PCA to define a framework for multiattribute analysis to understand which seismic attributes are significant for unsupervised learning. In their study, the combination of attributes determined by PCA is used as input to SOM to identify geologic patterns and to define stratigraphy, seismic facies, and direct hydrocarbon indicators. Zhao et al. (2018) build on these ideas and suggest a weight matrix computed from the skewness and kurtosis of attribute histograms to improve SOM learning.

In general, attribute selection in unsupervised learning relies on the data distribution of the input attributes and the correlation between input attributes. Supervised learning maps a relationship between input attributes and output using an interpreter-defined training data set. Several supervised learning studies introduced attribute selection methods, also known as feature selection or variable selection to reduce dimensionality (Jain and Zongker, 1997; Chandrashekar and Sahin, 2014). We present multiple strategies to select appropriate attributes for seismic facies classification with a case study. Our goals are to provide a good classification model in terms of validation accuracy, to avoid overfitting, and to reduce the computation and memory requirements needed for generating seismic attributes.

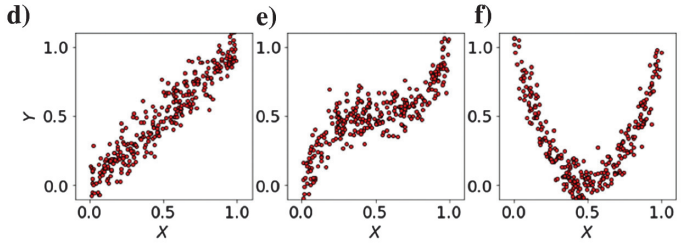
A desirable attribute subset might be built by detecting relevant attributes and discarding the irrelevant ones (Sánchez-Maroño et al., 2007). Although relevant attributes are those that are highly correlated with the output classes, redundant attributes are highly correlated with each other. Barnes (2007) suggests that there are many redundant and useless attributes that breed confusion in seismic interpretation; we argue that these attributes also pose problems in machine-learning classification.

To avoid building an unnecessarily complex model, we evaluate several attribute selection algorithms to maximize relevance and minimize redundancy to build an efficient subset of attributes for supervised facies classification analysis. Attribute selection methods can be classified into three groups: (1) a filter method that uses a correlation or dependency measure, (2) a wrapper method that applies a predictive model to evaluate the performance of an attribute subset, and (3) an embedded method, which measures the attribute importance during the training process. Because multiple attributes are analyzed simultaneously in the test, we consider our attribute selection algorithm to be a multivariate algorithm.

We compare the three types of attribute selection algorithms to build an efficient subset to differentiate seismic facies in a Gulf of Mexico survey. We generate 20 attributes from amplitude, instantaneous, geometric, texture, and spectral categories. The aim of the case study is to classify the specific facies based on patterns from a labeled training data set. We define the target classes of training data as being the facies corresponding to salt diapirs, MTDs, and conformal reflectors, which are created from manual geologic and stratigraphic interpretation. Correlations between attributes and correlations between attributes and output classes are analyzed using different measures to investigate



	a) Linear monotonic	b) Nonlinear monotonic	c) Nonlinear nonmonotonic
Correlation between X and Y	Pearson corr.=1.00 Rank corr.=1.00 MI=1.00 Distance corr.=1.00	Pearson corr.=0.92 Rank corr.=1.00 MI=1.00 Distance corr.=0.84	Pearson corr.=0.05 Rank corr.=−0.01 MI=1.00 Distance corr.=0.26
Regression score between X and Y	Linear Bayesian = 1.00 DNN = 1.00 RF = 1.00 SVM = 1.00	Linear Bayesian = 0.84 DNN = 1.00 RF = 1.00 SVM = 1.00	Linear Bayesian = -0.04 DNN = 1.00 RF = 1.00 SVM = 1.00



	d)	e)	f)
Correlation between X and Y	Pearson corr.=0.95 Rank corr.=0.95 MI=0.89 Distance corr.=0.88	Pearson corr.=0.84 Rank corr.=0.83 MI=0.62 Distance corr.=0.65	Pearson corr.=0.02 Rank corr.=−0.01 MI=0.87 Distance corr.=0.23
Regression score between X and Y	Linear Bayesian = 0.89 DNN = 0.89 RF = 0.84 SVM = 0.89	Linear Bayesian = 0.68 DNN = 0.81 RF = 0.74 SVM = 0.81	Linear Bayesian = -0.02 DNN = 0.90 RF = 0.86 SVM = 0.90

Figure 1. Different types of relationship between variables X and Y and their correlation coefficients and regression score. Each scatterplot describes a different relationship between X and Y: (a and c) linear and monotonic relationships, (b and e) nonlinear, monotonic relationship, and (c and f) nonlinear, nonmonotonic relationships. Gaussian noise of 10% has been added to variable Y in (d-f). Coefficients are computed using Pearson, rank, MI, and distance correlation methods. A regression score is computed for the linear Bayesian, NN, RF, and SVM repressor predictive algorithms. The best hyperparameters for each model were obtained using a grid-search algorithm.

the relevance and redundancy of each seismic attribute. The selected attributes are tested using a random forest (RF) algorithm, and the classification results are discussed. We also apply our workflow to the Barnett Shale play in the Fort Worth Basin to differentiate shale and limestone facies using inverted physical properties as input attributes. The output class is labeled based on stratigraphic interpretation aided by adjacent wireline logs. The classification results using different attribute subsets are discussed.

Correlation measures to maximize relevance and minimize redundancy

Finding an optimal subset can be achieved by maximizing the relevance between attributes and output classes, while minimizing redundancy among attributes (Yu and Liu, 2004; Peng et al., 2005). To maximize relevance, attributes that are highly correlated with output classes are selected. On the other hand, redundancy is caused by attributes that are highly correlated to each other. Thus, measuring and analyzing the correlation between attributes and classes, or correlations between attributes are prioritized to evaluate the performance of each attribute subset. Several correlation measures can be used in the feature selection. We examine Pearson's correlation (Pearson, 1894), rank correlation (Spearman, 1904), mutual information (MI) (Shannon and Weaver, 1949; Cover and Thomas, 1991), and distance correlation (Székely et al., 2007). Refer to Appendix A for a mathematical description. Pearson's correlation (Pearson, 1894) is the most common measure, and it detects only a linear relationship between two random variables. Spearman's rank correlation (Spearman, 1904) measures the tendency of a positive or negative relation, without requiring the increase or decrease to be explained by a linear relationship. Figure 1 illustrates different types of relationships between variables X and Y. Four types of correlation measure are able to detect the linear relationship (Figure 1a and 1d). In Figure 1b and 1e, the rank correlation has a higher coefficient value than Pearson's correlation because rank correlation is able to detect nonlinear positive relationships, whereas Pearson's correlation is not. In addition, dependence among attributes is not always linear. MI (Shannon and Weaver, 1949; Cover and Thomas, 1991) and distance correlation (Székely et al., 2007) detect nonlinear and nonmonotonic relationships. In Figure 1c and 1f, the Pearson's and rank correlation coefficients are approximately zero, which indicates that these two correlation measures do not detect nonlinear and nonmonotonic relationship.

In terms of dependence between an input attribute and an output class, it is also important to identify each predictive model's ability to map a nonlinear relationship between an attribute and a class. Even though the attribute is a powerful variable, which can have a high correlation with class, some predictive models may degrade prediction accuracy if the model cannot properly map the relationship. Figure 1 describes four types of predictive models: linear Bayesian, neural network (NN), RF, and support vector machine (SVM) and

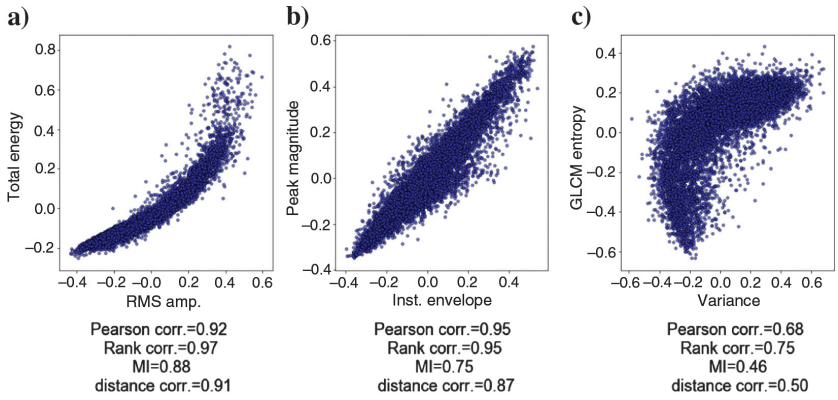


Figure 2. Relations between different attribute pairs (a) total energy versus rms amplitude, (b) peak magnitude versus instantaneous envelope, and (c) GLCM entropy versus variance. Correlation coefficients are computed using Pearson, rank, MI, and distance measures.

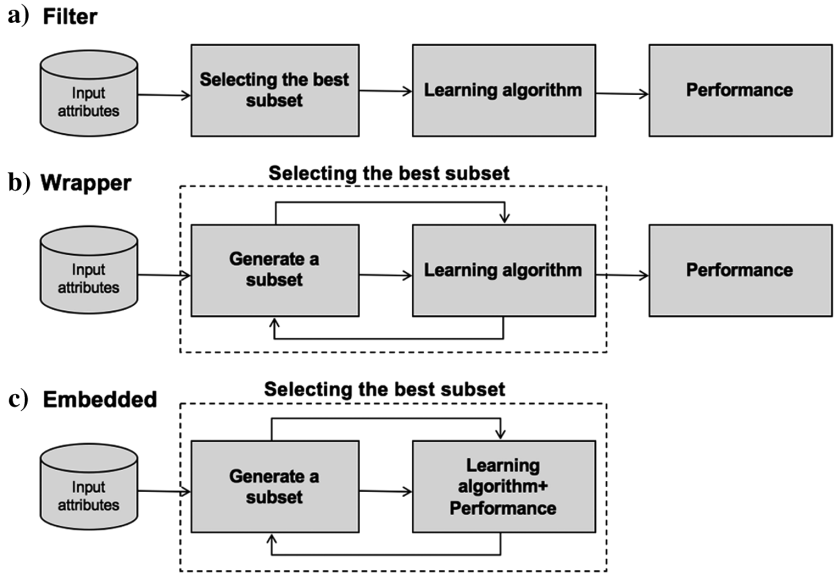


Figure 3. Schematic diagram summarizing the steps from the (a) filter, (b) wrapper, and (c) embedded attribute subset selection workflows. Note that there is no feedback in the filter workflow. The examples of each method and their mathematical description are given in Appendix B.

their ability to map input and output using regression methods. Using data points described in each plot, five-fold cross validation is applied. The hyper parameters for each predictive model were selected based on grid searches that give the best validation score. Linear-Bayesian models are not able to map a nonlinear relationship that gives an accuracy of 0.84 in monotonic case (Figure 1b) and −0.04 in a nonmonotonic case (Figure 1c). Except for linear-Bayesian, the other three models map the input and output with high accuracy (1.0) when noise is not added (Figure 1b and 1c). We select our test predictive model to be NN, RF, and SVM for our case study because they are able to map the nonlinear relationship appropriately. The noise in the signal can affect the correlation because it gives more uncertainty in predicting the output class. The sensitivity to noise also differs with correlation measure. MI and distance correlation coefficients decrease more than the others, when 10% of Gaussian noise is added to variable Y as shown in Figure 1d–1f.

The correlation measures are affected not only by nonlinearity but also by the covariance of two variables or of a variable with noise. Figure 2 shows the scatterplots of two attributes, which have relatively high correlations. The rms amplitude and total energy in Figure 2a have a positive, monotonic relationship because the energy attribute is equivalent to the square of the amplitude. These two attributes exhibit a high rank correlation coefficient (0.97). The relationship between peak magnitude and instantaneous envelope (Figure 2b) is linear, and it has a higher Pearson's coefficient than the other two scatterplots. Figure 2c describes correlation between entropy computed from gray-level cooccurrence matrix (GLCM) and variance. MI and distance correlation can detect nonlinear relationships, but their values for GLCM entropy and variance are lower in Figure 2c because their entropy is high.

In addition, we also test the analysis of variance (ANOVA) to determine which attributes are significant to differentiate output classes. ANOVA is an analysis tool that splits the variability found in a data set into systematic factors and random factors. If the variation can be explained from systematic factors, then the variable is significant in distinguishing classes.

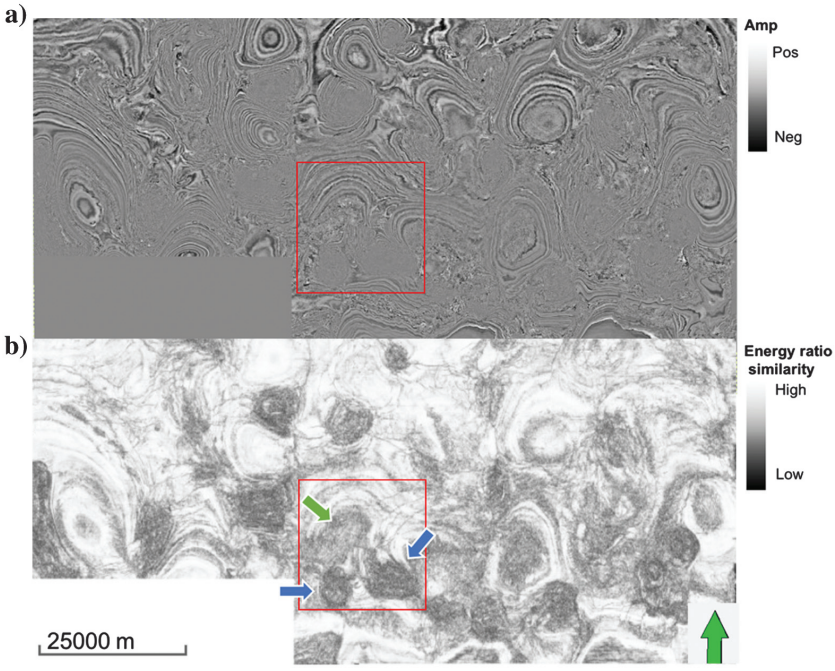


Figure 4. Time slice through the (a) seismic amplitude and (b) energy ratio similarity attribute. The red box indicates the volume in which the training data are sampled. The green arrow indicates MTDs, whereas the blue arrows indicate salt diapirs, both of which exhibit a low value of similarity.

Table 1. The seismic attribute families and the 20 specific attributes used to classify conformal sediments, salt, and MTDs.

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

Attribute selection algorithms: filters, wrappers, and embedded methods

The goal of attribute selection is to differentiate seismic facies effectively with an optimal combination of different attributes. To choose an optimal subset, the relationship between attributes as well as their relevance to the output should be analyzed in a multivariate manner. To measure redundancy is simple when attributes are perfectly correlated. If two attributes are perfectly correlated, then adding them does not provide additional information. Guyon and Elisseeff (2003), however, suggest that if two variables are highly correlated, then they have a possibility to complement each other. In addition, two variables that are not relevant by themselves can be useful when they are used together. Selecting attributes when considering relevance and redundancy together can be a complicated problem, but if the attribute selection algorithms are developed in a multivariate manner, they can be applied for attribute selection as well.

In supervised classification, there are three major approaches to select attributes in a multivariate manner: filters, wrappers, and embedded methods. Figure 3 describes the mechanisms of these methods. Filter methods use a suitable measure or ranking criterion, such as correlation or MI to select attributes. Relief (Kira and Rendell, 1992) is a distance-based filter algorithm that evaluates attributes according to feature value differences between nearest-neighbor instance pairs. ReliefF (Kononenko, 1994), an updated Relief algorithm, can deal with multiclass problems, and it is more robust to incomplete and noisy data. Correlation-based feature selection (CFS) (Hall, 1999) is an algorithm based on a heuristic evaluation function, which is calculated from attribute-class and attribute-attribute correlations. The fast correlation-based filter (FCBF) algorithm (Yu and Liu, 2003) is also a correlation-based measure but is designed for high-dimensional data. Filter methods are computationally less expensive than the wrapper method, which requires computation of a classification model.

Wrapper methods use a classification model to select the attribute subset. Wrapper methods require greater computational resources but provide better performance in that they maximize the classification accuracy. The sequential forward selection (SFS) algorithm (Kittler, 1978), for example, starts with an empty subset and adds

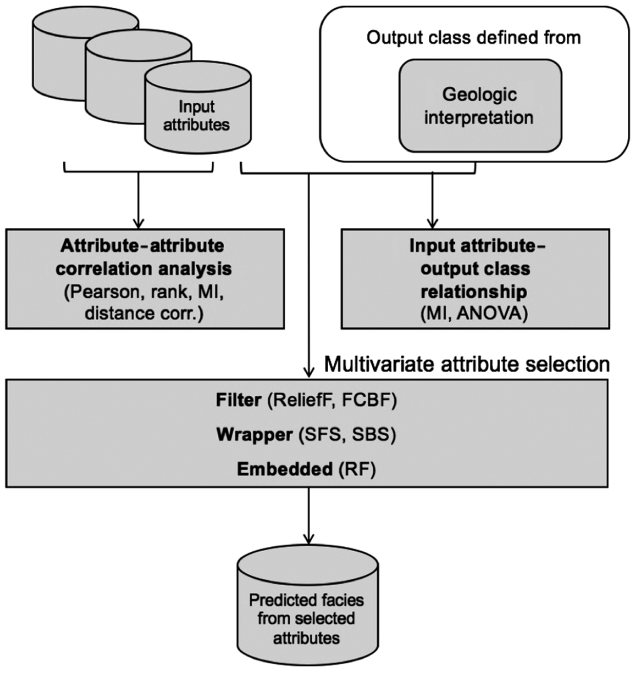


Figure 5. The workflow to select the best subset of attributes based on geologic relevance as well as attribute-to-attribute redundancy using three types of multivariate approaches.

Table 2. Attribute-to-attribute correlation analysis using Pearson, rank, MI, and distance correlations.

Attribute – attribute correlation analysis	
Correlation measures	**Attributes highly correlated with the other attributes (corr. coeff. >0.6)**
Pearson correlation	GLCM entropy - GLCM homogeneity (-1.0)
	Instantaneous envelope - Peak magnitude (0.96)
	RMS amplitude - Instantaneous envelope (0.93)
	RMS amplitude - Peak magnitude (0.90)
	RMS amplitude - Total energy (0.90)
	Total energy - Instantaneous envelope (0.84)
	Total energy -Peak magnitude (0.83)
	Instantaneous phase - Relative acoustic impedance (0.73)
	GLCM entropy - Chaos (0.71)
	GLCM entropy - Variance (0.70)
	Instantaneous frequency - Peak frequency (0.62)
Rank correlation	GLCM entropy - GLCM homogeneity (-1.0)
	RMS amplitude - Total energy (0.99)
	Instantaneous envelope - Peak magnitude (0.95)
	RMS amplitude - Instantaneous envelope (0.94)
	RMS amplitude - Total energy (0.93)
	RMS amplitude - Peak magnitude (0.92)
	Total energy - Peak magnitude (0.82)
	Instantaneous phase - Relative acoustic impedance (0.81)
	GLCM entropy - Variance (0.80)
	GLCM entropy - Chaos (0.71)
	Instantaneous frequency - Peak frequency (0.64)
	Instantaneous envelope - GLCM homogeneity (0.62)
Mutual information	RMS amplitude - Total energy (0.9)
	GLCM entropy - GLCM homogeneity (0.85)
	Instantaneous envelope - Peak magnitude (0.74)
	RMS amplitude - Instantaneous envelope (0.72)
	Total energy - Instantaneous envelope (0.71)
	RMS amplitude - Peak magnitude (0.69)
	Total energy - Peak magnitude (0.68)
Distance correlation	GLCM entropy - GLCM homogeneity (0.97)
	RMS amplitude - Total energy (0.92)
	Instantaneous envelope - Peak magnitude (0.87)
	RMS amplitude - Instantaneous envelope (0.83)
	RMS amplitude - Peak magnitude (0.78)
	Total energy - Instantaneous envelope (0.78)
	Total energy - Peak magnitude (0.74)

Note: Attribute pairs exhibiting a high correlation (correlation coefficient >0.6) are ranked in descending order. Pairs with a yellow background are the amplitude attributes or the attributes that are highly correlated with the amplitude attributes. Pairs with the green background are the texture attributes or the attributes highly correlated with the texture attributes.

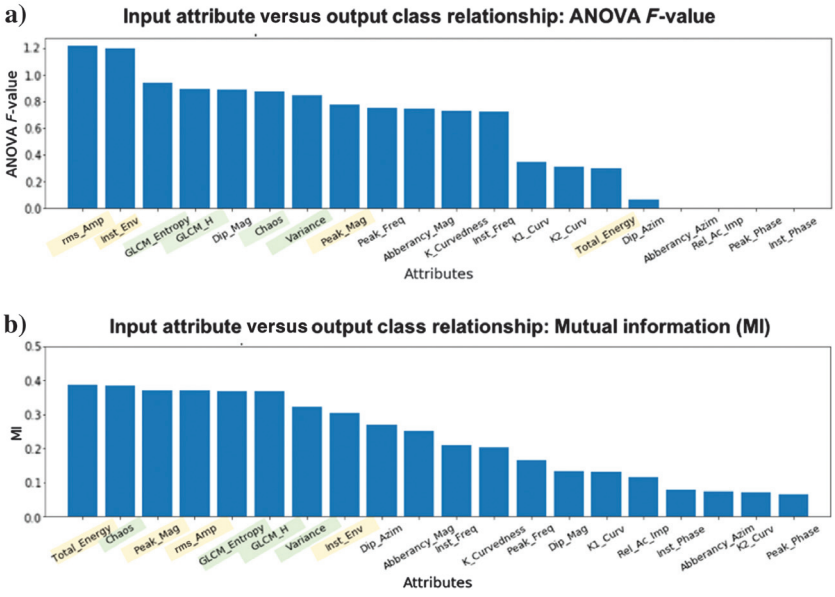


Figure 6. Relationships between a single input attribute and the desired output classes using (a) ANOVA *F*-value and (b) MI. Both analyses show that amplitude and texture type attributes are important variables for classification. (The yellow background indicates the amplitude attributes or the attributes highly correlated with amplitude attributes. The green background indicates texture attributes or the attributes highly correlated with texture attributes.) Unless the prediction is restricted to a specific horizon, the phase varies between 180° and +180° with increasing time, and it is poorly correlated to output class. Examining Figure 4, it is clear that the azimuth of the reflector dip, faults, and flexures for this data set also varies between 180° and +180° and it is not correlated to any one facies.

Table 3. Selected attribute subsets using filter (ReliefF and FCBF), wrapper (NN, RF, and SVM), and embedded (RF) methods.

Attribute selection algorithms			Ranking of attributes (10 highest ranked attributes)				
Filter	ReliefF		Peak phase	Peak frequency	Total energy	Relative acoustic impedance	Instantaneous envelope
			Instantaneous frequency	Instantaneous phase	Variance	Dip magnitude	Dip azimuth
	FCBF		Peak frequency	Peak phase	rms amplitude	Total energy	Relative acoustic impedance
			Instantaneous envelope	Instantaneous frequency	Instantaneous phase	Variance	Dip magnitude
Wrapper	SFS	NN	Total energy	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip azimuth	Dip magnitude	Aberrancy azimuth	Peak frequency	Most-positive curvature
		RF	Total energy	Chaos	Aberrancy magnitude	Instantaneous frequency	Dip azimuth
			Variance	Dip magnitude	Curvedness	Peak frequency	Aberrancy azimuth
		SVM	rms amplitude	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip magnitude	Peak frequency	Curvedness	Dip azimuth	Aberrancy azimuth
	SBS	NN	Total energy	Variance	Dip magnitude	Instantaneous frequency	Aberrancy magnitude
			Dip azimuth	Aberrancy azimuth	Chaos	Peak frequency	Most-positive curvature
		RF	Total energy	Chaos	Aberrancy magnitude	Peak frequency	Variance
			Dip azimuth	Dip magnitude	Most-positive curvature	Instantaneous frequency	Most-negative curvature
		SVM	rms amplitude	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip magnitude	Peak frequency	Most-negative curvature	Dip azimuth	Aberrancy azimuth
Embedded	RF		Total energy	Peak magnitude	Chaos	rms amplitude	Aberrancy magnitude
			GLCM homogeneity	GLCM entropy	Instantaneous envelope	Variance	Instantaneous frequency

Note: Each subset includes the 10 best attributes ranked in descending order (from left to right).

an attribute to the subset sequentially to yield the highest increase in score. Sequential backward selection (SBS), on the other hand, subtracts an attribute from a full subset sequentially whose elimination gives the lowest decrease in score.

Embedded methods implement attribute selection as a part of the training process of classification. In addition to having a low computational cost, embedded methods do not require a separate process for attribute selection. For instance, an RF classifier calculates the variable importance (Breiman, 2001; Liaw and Wiener, 2002) during training. Another embedded technique is to compute the weights of each attribute in the SVM classifier (Guyon et al., 2002) and logistic regression (Ma and Huang, 2005).

Application 1: Gulf of Mexico survey — Attribute selection to differentiate salt, mass transport deposits, and conformal reflectors

Seismic expression of salt and MTDs

Salt diapirs inherently have poor internal reflectivity and are easily overprinted by crossing coherent migration artifacts (Jones and Davison, 2014) in part due to their geometry and their higher P-wave velocities compared with surrounding strata. In general, the mismigrated noise gives rise to a relatively low amplitude, chaotic, and discontinuous seismic patterns that result in low coherence and high GLCM entropy inside the salt body. Therefore, texture attributes, such as GLCM (entropy, homogeneity, energy), are used to differentiate salt diapirs (Berthelot et al., 2013; Qi et al., 2016). Mass transport deposits (MTDs) are slumps, slides, and debris flows generated by gravity-controlled processes (Nelson et al., 2011). MTDs often show chaotic or highly disrupted seismic patterns with great internal complexity (Martinez, 2010). In general, the resulting attribute anomalies are high rms amplitudes and low coherence (Brown, 2011; Omosanya and Alves, 2013). The conformal reflectors around salt diapirs and MTDs show a relatively continuous seismic pattern that leads to high coherence and low to moderate values of GLCM entropy.

Methodology

The 3D marine seismic data were acquired in offshore Louisiana over an area of 3089 mi^2 (Qi et al., 2016). The poststack seismic volume includes 4367 inlines, 1594 crosslines, and 475 time samples with a sampling interval of 4 ms. Twenty seismic attributes in five categories were calculated from the seismic volume. The five attribute categories consist of amplitude, geometric, instantaneous, texture, and spectral attributes (Table 1). For supervised learning, we use a voxel-type training data set that is rendered from geologic and stratigraphic interpretation: Salt, MTDs, and conformal reflectors are interpreted inside the red box (751 inlines × 551 crosslines) as shown in Figure 4, and they are cropped as a 3D seismic volume using a polygon. From the cropped volume, 10,000 voxels were randomly selected for each facies and labeled for training

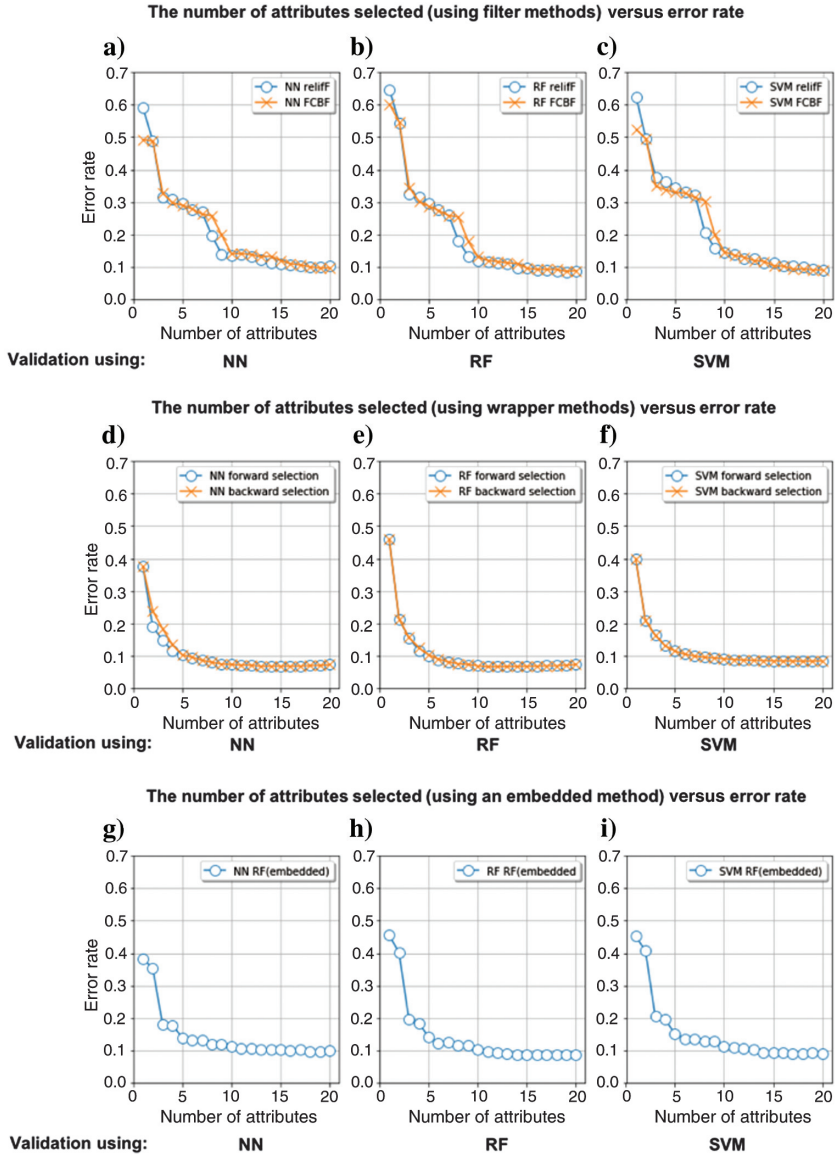


Figure 7. The number of attributes included in the attribute subset versus error rate. (a-c) Attributes in the subset were selected using filter methods (ReliefF and FCBF). (d-f) Attributes included in the subset were selected using wrapper methods (SFS and SBS). (g-i) Attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using the NN, RF, and SVM classifiers.

output (e.g., conformal reflectors: 0, salt diapirs: 1, and MTDs: 2). The training input data were then extracted from the 20 attribute volumes at the same voxel locations. Figure 5 summarizes the attribute selection workflow. First, we look into attribute-attribute correlations using four measures of correlation: Pearson, rank, MI, and distance correlation. These measures are valid to analyze the relationships between two continuous variables. To investigate the relationship between attributes and an output class, we used ANOVA and MI. Both metrics can provide information on how well a single attribute can differentiate classes and is relevant to each output class individually. Even if the correlation measures are not used to build the attribute subsets, correlation measures help to explain and evaluate the results from the attribute selection methods. We apply multivariate algorithms using three approaches: filter, wrapper, and embedded methods. Among several filter methods, ReliefF and FCBF are tested. For wrapper method, we applied SFS and SBS with three classifiers: NN, RF, and SVM. For the embedded method, the RF classifier is adopted, which produces a ranking of variables during the training process. For each method, we test the performance and evaluate the error rates of the attribute subset using the NN, RF, and SVM classifiers. We predict 3D facies using an RF classifier for the best attribute subset to test the validity of the model.

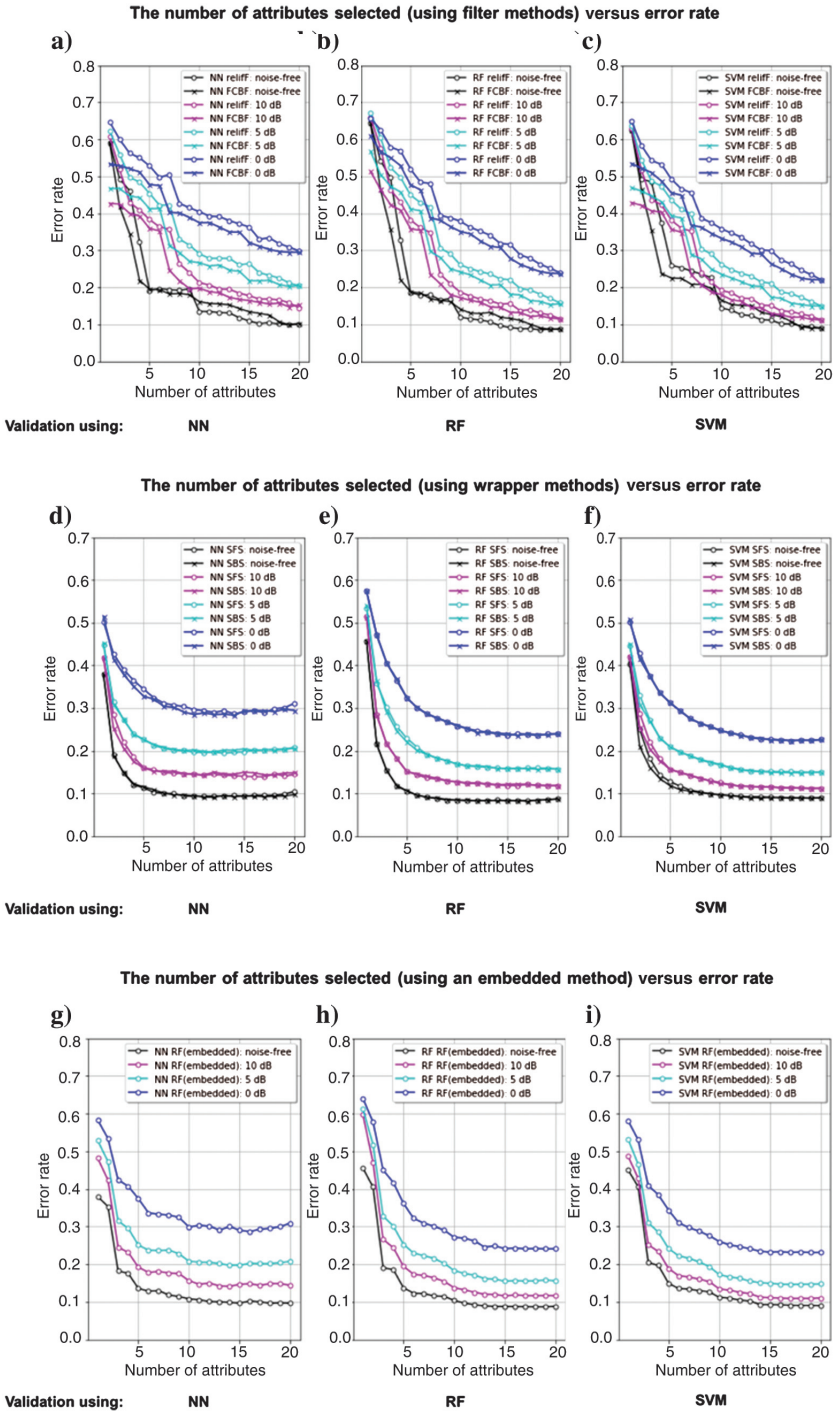


Figure 8. The number of attributes included in the attribute subset versus the error rate when Gaussian noise for different S/N levels is added (noise-free, 10, 5, and 0 dB) to attributes. (a-c) Attributes in the subset were selected using filter methods (ReliefF and FCBF). (d-f) Attributes included in the subset were selected using wrapper methods (SFS and SBS). (g-i) Attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using NN, RF, and SVM classifiers.

Results and discussion

Table 2 shows attribute-attribute correlation using the Pearson, rank, MI, and distance correlation measures. We note that correlation measures have different susceptibilities to nonlinearity, the presence of noise and outliers, and also whether or not the attributes are normally distributed. For instance, Pearson's correlation coefficient changes substantially compared to the rank correlation coefficient when an outlier is included (Mukaka, 2012). In the case of MI, the response of one variable is due to stimuli and noise (Shannon and Weaver, 1949). Figure 1d–1f shows that the presence of noise decreases MI and the distance correlation coefficient significantly. The GLCM homogeneity and entropy are perfectly anticorrelated when measured with the Pearson, rank, and distance correlation, suggesting that we can select only one of them for the subset to avoid redundancy. Amplitude attributes such as the rms amplitude and total energy are highly correlated (correlation coefficient >0.9). In addition,

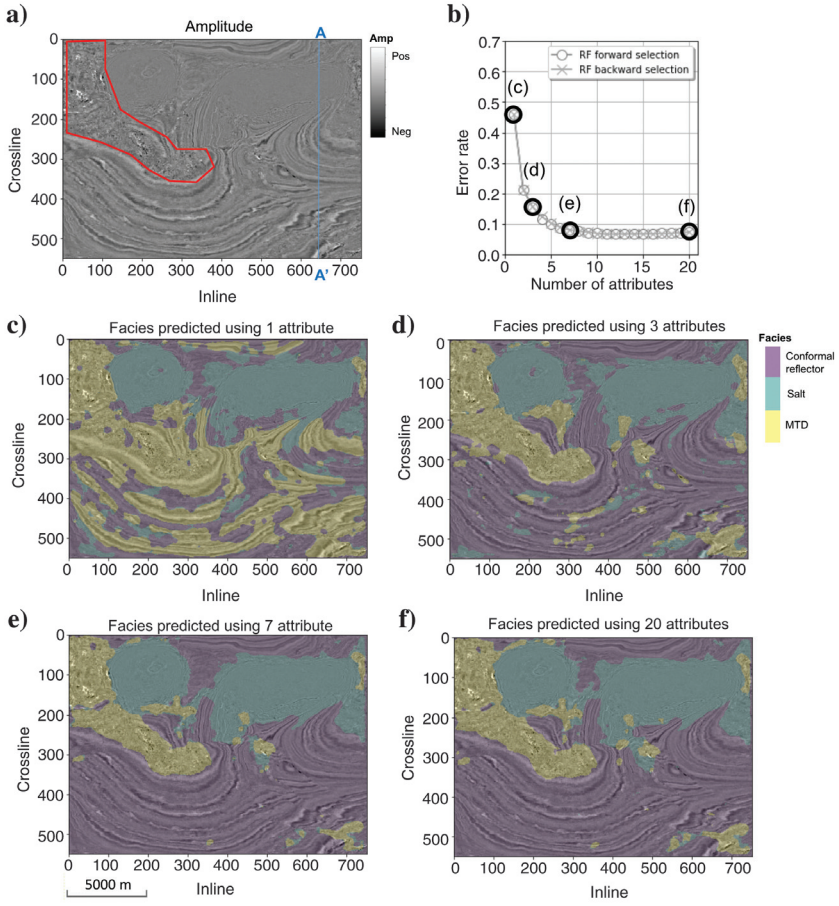


Figure 9. (a) A representative time slice at $t = 1.1$ s through amplitude and (b) error rate with respect to the number of attributes in the subset selected by the wrapper method (RF) using training data. Facies predicted using (c) the highest-ranked attribute, (d) top three highest-ranked attributes, (e) top seven highest-ranked attributes selected by the wrapper method (RF), and (f) all 20 attributes. The red polygon in (a) is a human-delineated MTD.

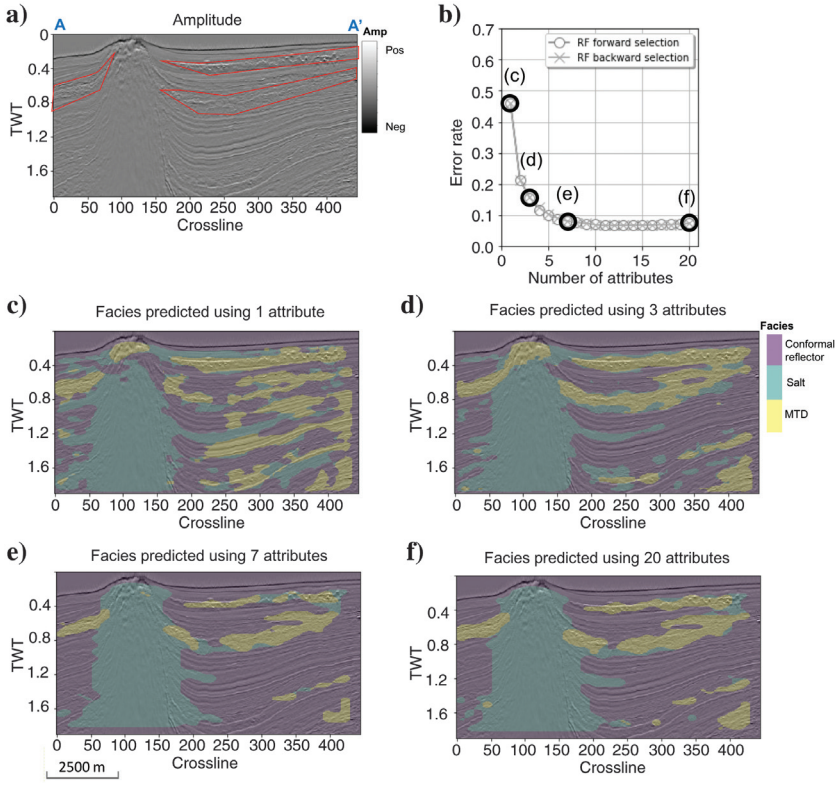


Figure 10. (a) A representative vertical slice along line AA′ through amplitude and (b) error rate with respect to the number of attributes in the subset, which is selected by the wrapper method (RF) using training data. Facies predicted using (c) the highest-ranked attribute, (d) the top three highest-ranked attributes, (e) the top seven highest-ranked attributes selected by the wrapper method (RF), and (f) all 20 attributes. The three red polygons in (a) are human-interpreted MTDs.

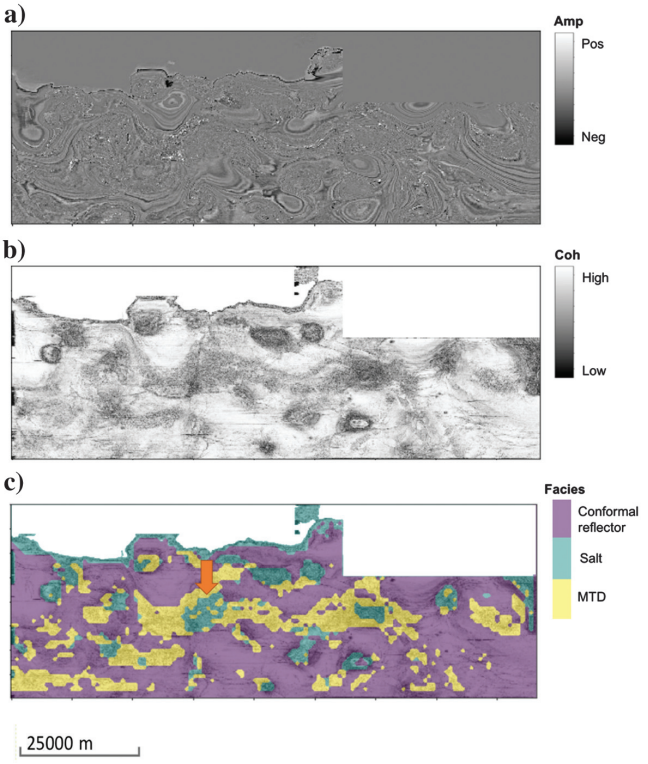


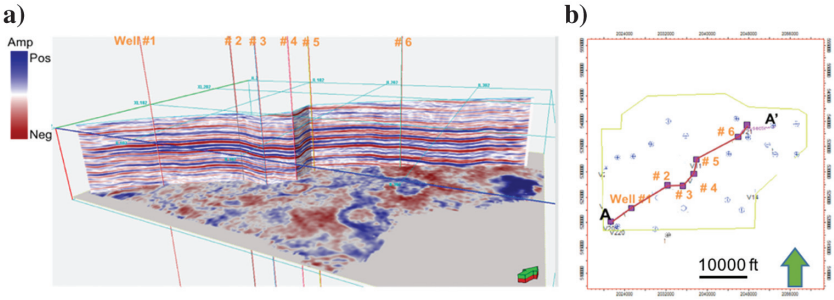
Figure 11. A time slice at $t = 0.612$ s through the (a) amplitude, (b) coherence, and (c) facies predicted using seven highest-ranked attributes using the wrapper method (RF). The arrow in (c) indicates the area where MTDs are misclassified as salt.

both of the amplitude attributes are highly correlated with instantaneous envelope and peak magnitude. Because MI is more sensitive to noise or the distribution of data points, the MI coefficients are lower than those of other measures.

The attribute-to-class relationship is analyzed using ANOVA and MI (Figure 6). Both methods show that the amplitude family of attributes (e.g., rms amplitude, total energy, instantaneous envelope, and peak magnitude) are relevant to the output classes. In addition, texture attributes (GLCM and chaos) are strongly related to training classes. Recall that ANOVA is based on a linear model, whereas MI is based on a nonlinear model. MI shows that the output classes have high dependence with total energy and rms amplitude. On the other hand, the ANOVA model indicated rms amplitude exhibits a high F-value. Although ANOVA and MI tell us which attributes can better differentiate the facies of interest, attributes that are selected by both methods can have redundancy. For instance, the GLCM entropy and GLCM homogeneity are highly ranked in ANOVA and MI (Figure 6), which shows that they are powerful variables for classification. However, we need only select one of two attributes for the subset. Because the two attributes are perfectly anticorrelated, this indicates that using both is redundant.

To take into account relevance and redundancy, we test several attribute selection algorithms to build the attribute subsets. The 10 highest ranked attributes

Figure 12. (a) The vertival slice along AA′ and the representative time slice through the seismic amplitude and (b) map of the seismic survey and location of the wells used for data training.



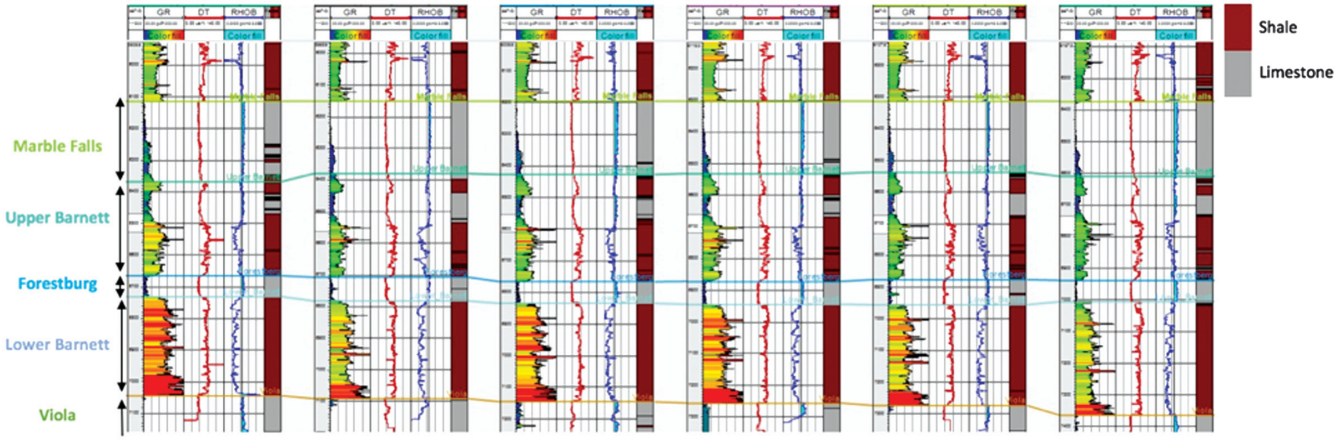


Figure 13. Well logs through the Barnett Shale showing the relevant section (Marble Falls — Upper Barnett — Forestburg — Lower Barnett — Viola). The section is flattened on the top Marble Falls. Wireline log data include gamma ray, P-sonic, and bulk density. Facies were estimated based on each log set. The gray color represents limestone, and brown is shale.

obtained from each attribute selection algorithm are shown in Table 3. Algorithms belonging to the same categories (e.g., the ReliefF and FCBF algorithms of the filter method and six algorithms of the wrapper method) show similar attribute rankings. However, the filter and the wrapper algorithms yield quite different attribute subsets. Wrapper methods select relevant attributes, according to input-to-output dependence. At the same time, wrapper methods more efficiently reject redundant attributes. For instance, when the total energy is chosen in the subset, the wrapper algorithm rejects the rms amplitude and vice versa. RF variable selection is an example of an embedded method that tends to choose important attributes but also includes redundant attributes. For instance, the subset has total energy and peak magnitude ranked close together, whereas GLCM homogeneity and entropy are ranked close as well. Figure 7 shows the error rate of the attribute subsets selected using the filter, wrapper, and embedded methods. A fivefold cross validation is implemented to compute the accuracy score and error rate when each attribute subset is applied. Input attributes were split into five groups randomly, one group was used as the test (validation) data set, while the other four groups were used as training data sets. The cross validation process is repeated five times, and the average value of accuracy is used to compute the error rate. Wrapper methods reduce the error rate with a small number of attributes compared to other methods because the methods are based on the performance of the predictive model.

To understand how noise in the data set affects classification performance, Gaussian noise with different signal-to-noise ratios (S/Ns) (noise free, 10, 5, and 0 dB) were added to attributes of the training data set (Figure 8). We measure the S/N as a ratio of signal power compared to noise power in dB. The higher level of noise in the data generally degrades the classification accuracy. One key point of observation is that using a larger number of attributes significantly reduces the error rate in the case of noisy data. Especially, the RF and SVM wrapper models in Figure 8e and 8f show that the error rate of 0 dB data decreases substantially as the number of attributes increases. This implies that if the data are contaminated with noise, using other attributes together can improve classification.

We tested subsets using an RF classifier with differing numbers of attributes to differentiate the salt, MTDs, and conformal reflectors in the same seismic volume that we used for the training set (Figures 9, 10, and 11). The error rate with respect to the number of attributes in each subset is computed from training data and is shown in Figure 9b. The subset consisting of just the first highest-ranked attribute does not differentiate MTDs and conformal reflectors (Figure 9c). The three highest-ranked attributes distinguish MTDs and conformal reflectors better than the one attribute subset (Figure 9d). However, some parts of the conformal reflectors are misclassified into salt and MTD. The top seven highest-ranked attributes differentiate the three facies as effectively as 20, the full set of attributes. The subset with the top seven highest-ranked attributes include relevant attributes that map different geology while also avoiding redundancy. Figure 11 shows the predicted facies within the entire seismic volume. The salt domes that show high coherence values in Figure 11b are correctly predicted as salt facies in Figure 11c. A limitation of this classification is that some of the MTD facies are misclassified as salt because both facies are highly discontinuous and have low coherence. In addition, other discontinuous geologic features, such as faults, are misclassified as MTDs.

Application 2: Barnett Shale Play in the Fort Worth Basin — Attribute selection to differentiate limestone and shale facies

We test our workflow on the Barnett Shale to classify limestone and shale facies that are dominant in the play. In the survey area, the Barnett Shale is separated into upper and lower shale units by a thin Forestburg

Table 4. Attribute-to-attribute correlation analysis using MI.

Attribute-attribute correlation analysis	
Correlation measures	Attributes highly correlated with the other attributes (correlation coefficient >0.6)
MI	Lambda-Lambda rho (0.96)
	Mu rho-Mu (0.93)
	I_S-Mu (0.93)
	Young's modulus-Mu (0.89)
	V_S-Mu (0.88)
	V_S-Young's modulus (0.87)
	Mu rho-Young's modulus (0.83)
	I_S-Young's modulus (0.83)
	I_P-Young's modulus (0.79)
	I_P-V_P (0.79)
	I_S-V_S (0.75)
	Mu rho-V_S (0.75)
	V_P-Lambda rho (0.74)
	V_P-Lambda (0.72)
	V_P/V_S-Lambda (0.69)
	Poisson's ratio-Lambda (0.69)
	V_P-Young's modulus (0.65)
	I_P-Mu (0.64)
	I_P-V_S (0.63)
	I_P-Lambda rho (0.63)
	I_P-I_S (0.62)
	I_P-Mu rho (0.62)

Note: Attribute pairs exhibiting high correlation (correlation coefficient >0.6) are ranked in descending order.

Limestone. The Barnett Shale, which is relatively brittle and acts as the reservoir, lies between the Marble Falls and the Viola Limestones that are more ductile (Perez and Marfurt, 2014; Li et al., 2016; Verma et al., 2016). In this example, defining the output classes of the training data is aided by wireline logs and stratigraphic interpretation. A vertical slice through a seismic line along six wells is described in Figure 12. Facies were estimated based on each log set including gamma ray, P-wave sonic, and bulk density. We defined the facies mainly based on gamma ray, with values of limestone ranging from 10 to 40 API, whereas those of shale range from 60 to 150 API. From a stratigraphic interpretation that was aided by well logs, we labeled data points adjacent

Table 5. Selected attribute subsets using filter (ReliefF and FCBF), wrapper (NN, RF, and SVM), and embedded (RF) methods.

Attribute selection algorithms			Ranking of attributes (10 highest ranked attributes)				
Filter	ReliefF		Lambda rho	Lambda	Mu	Poisson	Young
			Rho	V_P/V_S	V_S	V_P	MuRho
	FCBF		V_S	Mu	I_S	V_P	I_P
			Mu rho	Young	Lambda	LambdaRho	Rho
Wrapper	SFS	NN	Young	I_P	Rho	V_P/V_S	Lambda rho
			Mu rho	V_S	V_P	Poisson	I_S
		RF	Young	I_P	Rho	Lambda	Lambda rho
			I_S	V_P/V_S	V_P	Poisson	V_S
		SVM	Young	I_P	Mu	V_S	I_S
			V_P/V_S	Rho	Mu rho	Poisson	Lambda rho
	SBS	NN	I_S	Poisson	V_S	Lambda rho	Mu rho
			V_P/V_S	V_P	Young	Lambda	Mu
		RF	V_P	I_S	Rho	Lambda	V_S
			I_P	Mu rho	Lambda rho	Mu	Young
		SVM	I_P	I_S	V_S	Mu rho	Poisson
			V_P	Young	V_P/V_S	Lambda	Mu
Embedded	RF		Young	I_P	Mu	V_S	V_P
			Mu rho	I_S	Lambda rho	Lambda	Rho

Note: Each subset includes 10 best attributes ranked in descending order (from left to right).

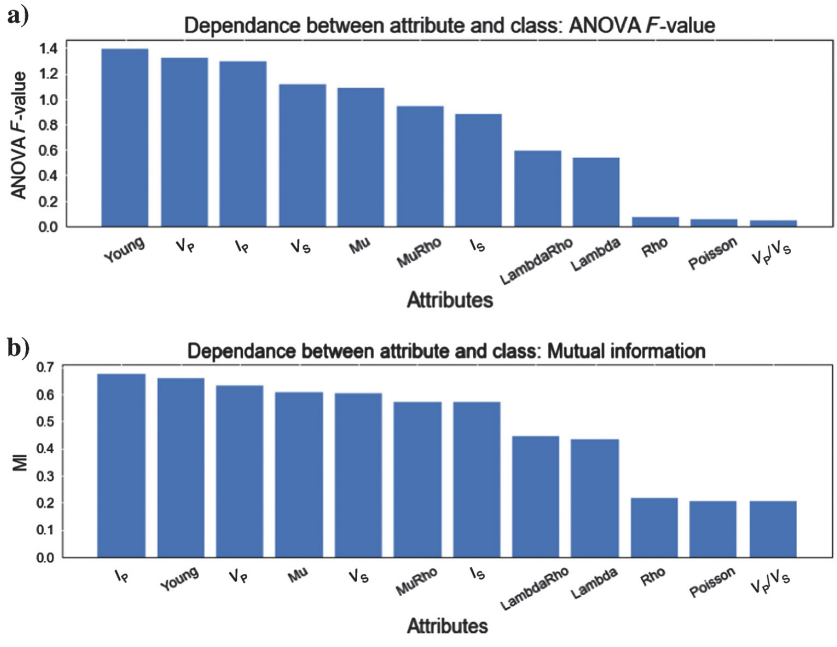


Figure 14. Relationships between a single input attribute and the desired output classes using (a) ANOVA *F*-value and (b) MI.

to each well log as limestone and shale, which is equivalent to the training output (Figure 13). The input attributes are comprised of 12 physical properties calculated from prestack seismic inversion: P-and S-impedances, P and S velocities, V_P/V_S, density, mu rho, Young's modulus, Poisson's ratio, mu, lambda, and lambda rho. The general workflow is the same as that of the first case study: attribute-attribute correlations and attribute-class correlations are analyzed, and attribute subsets are built using filter and wrapper algorithms. Among the four correlation measures, we opted for MI for the second case study because it is able to assess nonlinear relationships.

The attributes that describe physical properties inverted from seismic amplitude are highly correlated with each other because many of these physical properties can be calculated from other physical properties in the attribute set (Table 4). The elastic properties can especially be determined from two elastic moduli in the case of homogeneous isotropic media. In terms of attribute-class correlations, MI coefficients from seven attributes (P-impedance, Young's modulus, P velocity, mu, S velocity, mu rho, and S-impedance) are higher than 0.5 (Figure 14b). From the correlation analysis, we found that the attributes are highly correlated with each other, and many of the attributes are also highly correlated to the corresponding target class. The Young's modulus and P-impedance are highly ranked in the SFS of the wrapper methods (Table 5), and they decrease error rate efficiently (Figure 15d–15f). Because of the high redundancy in input attributes, the score in Figure 15 does not increase significantly after three components in each method.

Like the first case study, we tested subsets with differing numbers of attributes to differentiate shale and limestone in the same seismic volume that we used for the training set (Figure 16). Two thin lime layers are intervening in the Barnett Shale, which is interpreted in well-log section in Figure 13: a limestone layer in the Upper Barnett and the Forestburg Limestone between the Upper and Lower Barnett Shale. The subsets with the four highest-ranked attributes (Young's modulus, P-impedance, density, and Lambda) differentiate thin limestone layers between the Barnett Shale as effectively as 12 attributes (Figure 16).

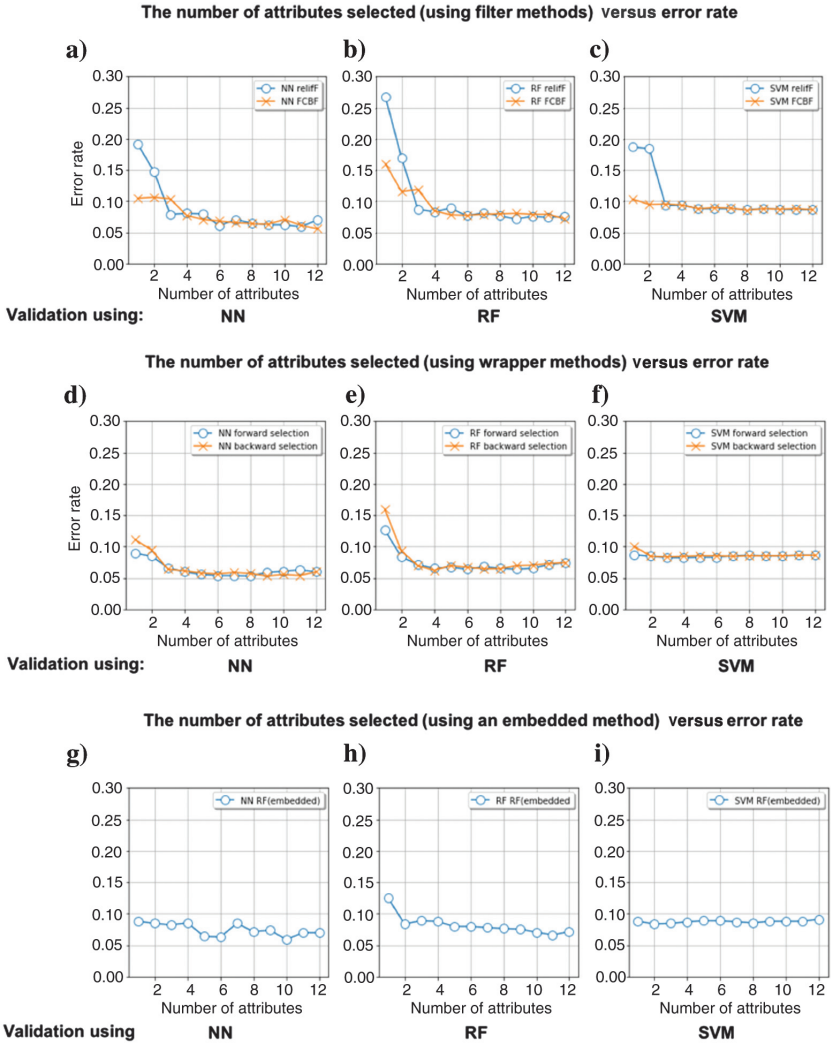


Figure 15. The number of attributes included in the attribute subset versus error rate. (a-c) Attributes in the subset were selected using filter methods (ReliefF and FCBF). (d-f) Attributes included in the subset were selected using wrapper methods (SFS and SBS). (g-i) Attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using NN, RF, and SVM classifiers.

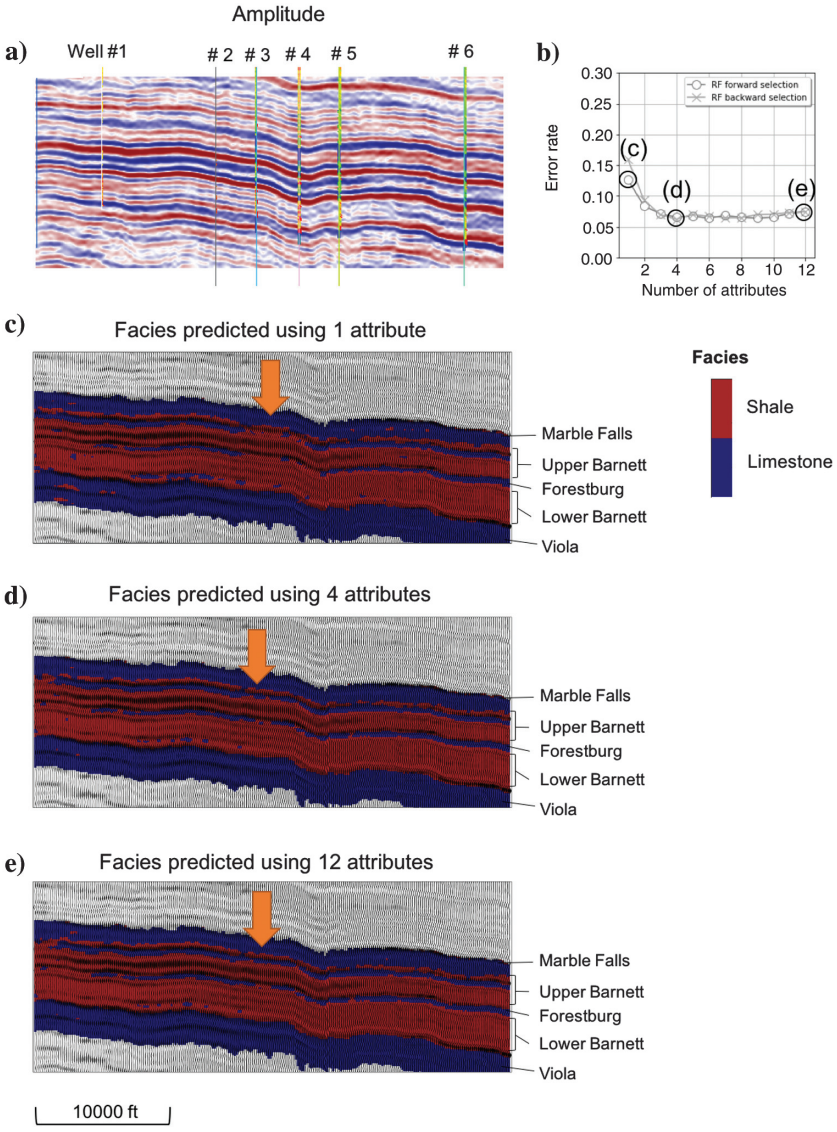


Figure 16. (a) A representative vertical slice through the amplitude and well logs and (b) the error rate with respect to the number of attributes in each subset, which is selected by the wrapper method (RF) using training data. Facies predicted using (c) the first highest-ranked attribute, (d) the top four highest-ranked attributes selected by the wrapper method (RF), and (e) all 12 attributes. Subsets with the top 4 highest-ranked attributes differentiate the thin limestone layers as effectively as all 12 attributes (the orange arrows).

Conclusion

Analyzing attribute-to-attribute dependence and attribute-to-class relationships helps to understand which attributes are redundant and which are relevant. However, a high correlation between the attributes does not always imply that attributes are redundant. We need to analyze all attributes together using a framework, which can quantitatively rank the attributes to build a subset. The multivariate attribute selection algorithms result in the subsets, which have smaller number of attributes but show good performance in differentiating salt and MTD facies from conformal reflectors. From a geologic point of view, it is challenging to define the depositional environments in the survey area into only three discrete classes. Turbidites, faults, overpressured shale, and seismic noise will be misclassified into one of the target classes. However, understanding each seismic attribute's characteristic is crucial to implement automated facies classification and to aid rendering of a seismic volume in which the interpreters target. Even though the case study is focused on mapping different facies in geology, the attribute selection algorithms can be applied to other supervised classification problems. For instance, the workflow can be applied to select physical properties and seismic attributes to yield reservoir properties such as porosity, permeability, and brittleness from input quantitative interpretation attributes.

Acknowledgments

The authors thank the sponsors of the OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support. We also thank Schlumberger for Petrel software license provided to The University of Oklahoma for research and education.

Data and materials availability

Data used are covered by a licensing agreement and cannot be shared with the general public.

Appendix A

Correlation measures

Pearson's product-moment correlation (Pearson, 1894)

A correlation measure that is most widely used is Pearson's product-moment coefficient. The correlation between two random variables X and Y is defined as

$$\mathbf{corr}(\mathbf{X},\mathbf{Y}) = \frac{\mathbf{cov}(\mathbf{X},\mathbf{Y})}{\boldsymbol{\sigma}_{\mathbf{x}}\boldsymbol{\sigma}_{\mathbf{y}}}, \tag{A-1}$$

where $\mathrm{cov}(X,Y)$ is the covariance between X and Y and σ_x and σ_y are the standard deviation of X and Y, respectively. The Pearson's coefficient describes the linear dependence between two variables. Among the scatterplots in Figure 1a–1c, only Figure 1a is perfectly correlated or anticorrelated in terms of Pearson's correlation.

Spearman's rank correlation (Spearman, 1904)

The Spearman's correlation coefficient is defined as Pearson's correlation between the ranked variables. A positive Spearman correlation corresponds to an

increasing monotonic trend, whereas a negative one corresponds to a decreasing monotonic trend between two random variables. The correlation assesses positive or negative relationships whether they are linear or not. According to the Spearman's definitions of correlation, two variables X and Y in Figure 1b are highly correlated even if the relationship is nonlinear.

Mutual influence (Shannon and Weaver, 1949; Cover and Thomas, 1991)

In information theory, the uncertainty involved in the value of a random variable is quantified as entropy. The Shannon entropy, which is a measure of the uncertainty of a random variable, is defined as

$$\mathbf{H} = -\sum_i \mathbf{p_i}\,\mathbf{log}(\mathbf{p_i}), \tag{A-2}$$

where $\mathbf{p_i}$ is the probability of occurrence of the ith possible value of the source symbol. Mutual influence (MI) measures the gain of information about one random variable by observing another. The MI of two discrete random variables X and Y is denoted by

$$\begin{aligned}\mathbf{I}(\mathbf{X};\mathbf{Y}) &= \mathrm{H}(\mathbf{X}) - \mathrm{H}(\mathbf{X}|\mathbf{Y})\\ &= \mathrm{H}(\mathbf{Y}) - \mathrm{H}(\mathbf{Y}|\mathbf{X}),\end{aligned} \tag{A-3}$$

where $\mathrm{H}(\mathbf{X})$ and $\mathrm{H}(\mathbf{Y})$ are the marginal entropies and $\mathrm{H}(\mathrm{X}|\mathrm{Y})$ and $\mathrm{H}(\mathrm{Y}|\mathrm{X})$ are the conditional entropies. Substituting equation A-2 to equation A-3 gives

$$\mathbf{I}(\mathbf{X};\mathbf{Y}) = \sum_{\mathbf{y}\in\mathbf{Y}}\sum_{\mathbf{x}\in\mathbf{X}} \mathrm{p}(\mathrm{x},\mathrm{y})\,\mathbf{log}\left(\frac{\mathrm{p}(\mathrm{x},\mathrm{y})}{\mathrm{p}(\mathrm{x})\mathrm{p}(\mathrm{y})}\right), \tag{A-4}$$

where $\mathrm{p}(\mathrm{x})$ and $\mathrm{p}(\mathrm{y})$ are the marginal probability functions and $\mathrm{p}(\mathrm{x},\mathrm{y})$ is the joint probability function.

Distance correlation (Székely et al., 2007)

The distance correlation of two random variables is defined as distance covariance divided by distance standard deviation. The distance correlation is denoted as

$$\mathrm{dCor}(\mathrm{X},\mathrm{Y}) = \frac{\mathrm{dCov}(\mathrm{X},\mathrm{Y})}{\sqrt{\mathrm{dVar}(\mathrm{X})\mathrm{dVar}(\mathrm{Y})}}, \tag{A-5}$$

where $\mathrm{dCov}(\mathrm{X},\mathrm{Y})$ is the distance covariance and $\mathrm{dVar}(\mathrm{X})$ and $\mathrm{dVar}(\mathrm{Y})$ are the distance variance of X and Y, respectively. In contrast to Pearson's covariance that is defined as the inner product of two centered vectors, the distance covariance is defined as the product of centered Euclidean distances $\mathrm{D}(\mathrm{x}_i,\mathrm{x}_j)$ and $\mathrm{D}(\mathrm{y}_i,\mathrm{y}_j)$ in arbitrary dimensions:

$$\mathrm{dCov}(\mathrm{X},\mathrm{Y}) = \frac{\mathbf{1}}{\mathbf{n^2}}\sum_{\mathbf{i}=1}^{\mathbf{n}}\sum_{\mathbf{j}=1}^{\mathbf{n}} \mathbf{D}(\mathbf{x_i},\mathbf{x_j})\cdot\mathbf{D}(\mathbf{y_i},\mathbf{y_j}), \tag{A-6}$$

where $\mathrm{x}\in\mathrm{X}$, $\mathrm{y}\in\mathrm{X}$, and n is the number of samples of X and Y. Distance correlation can detect nonlinear relationships, and its values are nonnegative.

Appendix B
Attribute selection methods

Filter

Relief and ReliefF

The Relief algorithm (Kira and Rendell, 1992) estimates attributes according to how well their values differentiate among the instances near to each other. Relief searches for its two nearest neighbors: one from the same class that is called the nearest hit and the other from a different class called the nearest miss. At each iteration, Relief estimates the weight vector W of a given attribute:

$$\begin{aligned}W_i = W_i &+ (x_i - \text{nearest miss}(x)_i)^2\\ &- (x_i - \text{nearest hit}(x)_i)^2,\end{aligned} \tag{B-1}$$

where x is an instance of randomly selected in training data. Attributes are selected if their average weight is greater than a threshold τ. ReliefF (Kononenko, 1994; Kononenko et al., 1997) improved Relief by estimating probabilities more reliably and extended the algorithm to handle noisy, incomplete, and multiclass data sets.

CFS and FCBF

CFS's feature subset evaluation function (Hall, 1999) is

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}, \tag{B-2}$$

where M_s is the heuristic merit of a feature subset containing k attributes, $\overline{r_{cf}}$ is the mean attribute-class correlation, and $\overline{r_{ff}}$ is the average attribute-attribute correlation. FCBF (Yu and Liu, 2003) starts with full set of features, uses symmetrical uncertainty to calculate dependences of features, and it finds the best subset using backward selection for high-dimensional data.

Wrapper

Sequential selection algorithms

The SFS algorithm (Kittler, 1978) starts from an empty set and sequentially adds the attributes that maximize the classification accuracy. The process is repeated until the required number of features are added. The SBS algorithm starts from the full set and sequentially removes the attribute that its removal gives the lowest decrease in classification performance. Sequential floating forward selection and sequential floating backward selection (Pudil et al., 1994) introduce an additional backtracking step that is more flexible than the naive SFS algorithm (Chandrashekar and Sahin, 2014).

Embedded

Attribute importance in tree-based methods (RF)

In the classification and regression tree algorithm (Breiman, 2001; Breiman, 2002), the best split is made using a Gini impurity at each internal node for prediction.

The Gini importance can be computed as byproduct during training process of tree-based predicted model, which is given by

$$\text{Gini importance } i_G(\theta) = \sum_T \sum_\tau \Delta i_\theta(\tau, \mathrm{T}), \tag{B-3}$$

where $\Delta i(\tau)$ is the node purity gain that is denoted as

$$\Delta i(\tau) = i(\tau) - p_l i(\tau_l) - p_r i(\tau_r). \tag{B-4}$$

Decrease in Gini impurity, $\Delta i(\tau)$ results from splitting the samples to left and right subnodes.

References

Barnes, A. E., 2007, Redundant and useless seismic attributes: Geophysics, **72**, no. 3, P33–P38, doi: 10.1190/1.2716717.

Berthelot, A., A. H. Solberg, and L. J. Gelius, 2013, Texture attributes for detection of salt: Journal of Applied Geophysics, **88**, 52–69, doi: 10.1016/j.jappgeo.2012.09.006.

Breiman, L., 2001, Random forests: Machine Learning, **45**, 5–32, doi: 10.1023/A:1010933404324.

Breiman, L., 2002, Manual on setting up, using, and understanding random forests v3. 1: Statistics Department University of California

Brown, A. R., 2011, Interpretation of three-dimensional seismic data: SEG and AAPG.

Chandrashekar, G., and F. Sahin, 2014, A survey on feature selection methods: Computers and Electrical Engineering, **40**, 16–28, doi: 10.1016/j.compeleceng.2013.11.024.

Chopra, S., and K. J. Marfurt, 2005, Seismic attributes — A historical perspective: Geophysics, **70**, no. 5, 3SO–28SO, doi: 10.1190/1.2098670.

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, doi: 10.1190/1.1623635.

Cover, T. M., and J. A. Thomas, 1991, Entropy, relative entropy and mutual information: Elements of Information Theory, **2**, 1–55.

Guyon, I., and A. Elisseeff, 2003, An introduction to variable and feature selection: Journal of Machine Learning Research, **3**, 1157–1182.

Guyon, I., J. Weston, S. Barnhill, and V. Vapnik, 2002, Gene selection for cancer classification using support vector machines: Machine Learning, **46**, 389–422.

Hall, M. A., 1999, Correlation-based feature selection for machine learning: Ph.D. thesis, University of Waikato.

Hughes, G., 1968, On the mean accuracy of statistical pattern recognizers: IEEE Transactions on Information Theory, **14**, 55–63, doi: 10.1109/TIT.1968.1054102.

Jain, A., and D. Zongker, 1997, Feature selection: Evaluation, application, and small sample performance: IEEE Transactions on Pattern Analysis and Machine Intelligence, **19**, 153–158, doi: 10.1109/34.574797.

Jones, I. F., and I. Davison, 2014, Seismic imaging in and around salt bodies: Interpretation, **2**, no. 4, SL1–SL20, doi: 10.1190/INT-2014-0033.1.

Kira, K., and L. A. Rendell, 1992, A practical approach to feature selection: International Conference on Machine Learning, 368–377.

Kittler, J., 1978, Feature set search algorithms: Pattern Recognition and Signal Processing, 41–60.

Kononenko, I., 1994, Estimating attributes: Analysis and extensions of RELIEF: European Conference on Machine Learning.

Kononenko, I., E. Šimec, and M. Robnik-Šikonja, 1997, Overcoming the myopia of inductive learning algorithms with RELIEFF: Applied Intelligence, **7**, 39–55, doi: 10.1023/A:1008280620621.

Li, F., S. Verma, H. Zhou, T. Zhao, and K. J. Marfurt, 2016, Seismic attenuation attributes with applications on conventional and unconventional reservoirs: Interpretation, **4**, no. 1, SB63–SB77, doi: 10.1190/INT-2015-0105.1.

Liaw, A., and M. Wiener, 2002, Classification and regression by Random Forest: R News, **2**, 18–22.

Ma, S., and J. Huang, 2005, Regularized ROC method for disease classification and biomarker selection with microarray data: Bioinformatics, **21**, 4356–4362.

Marfurt, K. J., R. L. Kirlin, S. L. Farmer, and M. S. Bahorich, 1998, 3-D seismic attributes using a semblance-based coherency algorithm: Geophysics, **63**, 1150–1165, doi: 10.1190/1.1444415.

Martinez, F. J., 2010, 3D seismic interpretation of mass transport deposits: Implications for basin analysis and geohazard evaluation, *in* D. C. Mosher, R. C. Shipp, L. Moscardelli, J. D. Chaytor, C. D. P. Baxter, H. J. Lee, and R. Urgeles, eds., Submarine mass movements and their consequences: Springer, 553–568.

Mukaka, M. M., 2012, A guide to appropriate use of correlation coefficient in medical research: Malawi Medical Journal, **24**, 69–71.

Nelson, C. H., C. A. Escutia, J. E. Damuth, and D. C. Twichell, 2011, Interplay of mass-transport and turbidite-system deposits in different active tectonic and passive continental margin settings: External and local controlling factors: Sedimentary Geology, **96**, 39–66.

Omosanya, K. O., and T. M. Alves, 2013, A 3-dimensional seismic method to assess the provenance of mass-transport deposits (MTDs) on salt-rich continental slopes (Espírito Santo Basin, SE Brazil): Marine and Petroleum Geology, **44**, 223–239, doi: 10.1016/j.marpetgeo.2013.02.006.

Pearson, K., 1894, Contributions to the mathematical theory of evolution: Philosophical Transactions A, **185**, 71–110, doi: 10.1098/rsta.1894.0003.

Peng, H., F. Long, and C. Ding, 2005, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy: IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 1226–1238, doi: 10.1109/TPAMI.2005.159.

Perez, R., and K. J. Marfurt, 2014, Mineralogy-based brittleness prediction from surface seismic data: Application to the Barnett Shale: Interpretation, **2**, no. 4, T255–T271, doi: 10.1190/INT-2013-0161.1.

Pudil, P., J. Novovičová, and J. Kittler, 1994, Floating search methods in feature selection: Pattern Recognition Letters, **15**, 1119–1125, doi: 10.1016/0167-8655(94)90127-9.

Qi, J., T. Lin, T. Zhao, F. Li, and K. J. Marfurt, 2016, Semisupervised multiattribute seismic facies analysis: Interpretation, **4**, no. 1, SB91–SB106, doi: 10.1190/INT-2015-0098.1.

Roden, R., T. Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **3**, no. 4, SAE59–SAE83, doi: 10.1190/INT-2015-0037.1.

Roy, A., M. Matos, and K. J. Marfurt, 2010, Automatic seismic facies classification with Kohonen self organizing maps — A tutorial: Geohorizons Journal of Society of Petroleum Geophysicists, **15-2**, 6–14.

Sánchez-Maroño, N., A. Alonso-Betanzos, and M. Tombilla-Sanromán, 2007, Filter methods for feature selection — A comparative study: International Conference on Intelligent Data Engineering and Automated Learning, 178–187.

Shannon, C., and W. Weaver, 1949, The mathematical theory of communication: University of Illinois Press.

Spearman, C., 1904, The proof and measurement of association between two things: American Journal of Psychology, **15**, 72–101, doi: 10.2307/1412159.

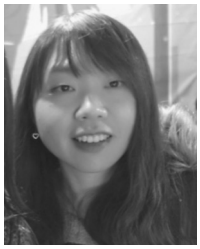
Székely, G. J., M. L. Rizzo, and N. K. Bakirov, 2007, Measuring and testing dependence by correlation of distances: The Annals of Statistics, **35**, 2769–2794, doi: 10.1214/009053607000000505.

Verma, S., T. Zhao, K. J. Marfurt, and D. Devegowda, 2016, Estimation of total organic carbon and brittleness volume: Interpretation, **4**, no. 3, T373–T385, doi: 10.1190/INT-2015-0166.1.

Yu, L., and H. Liu, 2003, Feature selection for high-dimensional data: A fast correlation-based filter solution: Proceedings of the 20th International Conference on Machine Learning, 856–863.

Yu, L., and H. Liu, 2004, Efficient feature selection via analysis of relevance and redundancy: Journal of Machine Learning Research, **5**, 1205–1224.

Zhao, T., F. Li, and K. J. Marfurt, 2018, Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights: Geophysics, **83**, no. 2, O31–O44, doi: 10.1190/geo2017-0192.1.



Yuji Kim received a B.S. in geology and an M.S. in exploration geophysics from Seoul National University, Seoul, South Korea. She is currently pursuing a Ph.D. in geophysics from the University of Oklahoma as a member of the Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium. Her research interests include developing and applying machine-learning techniques on seismic attributes development and signal processing.

Robert Hardisty received a B.S. and an M.S. in geophysics from the University of Oklahoma. As a former member of the AASPI consortium, he researched clustering techniques and their applications to seismic interpretation. He is currently working at Flywheel Energy, LLC as a geology data specialist providing data management solutions and analysis for asset development.

Kurt J. Marfurt is a research professor of geophysics at the University of Oklahoma, leads the Attribute-Assisted Seismic Processing and Interpretation consortium with the goal of developing and calibrating new seismic attributes to aid in seismic processing, seismic interpretation, and data integration using both interactive and machine-learning tools. His experience includes 23 years as an academician and 18 years in technology development at Amoco's Tulsa Research Center. He served as the 2006 EAGE/SEG and as the 2018 SEG Distinguished Short Course Instructor. He has taught continuing education short courses for SEG and AAPG since 2003. From 1984–2013, he has served as either an associate or assistant editor for GEOPHYSICS. In 2013 he joined the editorial board of the SEG/AAPG journal *Interpretation*, where he served as the Editor-in-Chief for 2016–2018, and now serves as Deputy Editor-in-Chief for 2019–2021.