

AVA classification as an unsupervised machine-learning problem

Ben B. Bougher and Felix J. Herrmann

Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

SUMMARY

Much of AVA analysis relies on characterizing background trends and anomalies in pre-stack seismic data. Analysts reduce a seismic section into a small number of these trends and anomalies, suggesting that a low-dimensional structure can be inferred from the data. We describe AVA-attribute characterization as an unsupervised-learning problem, where AVA classes are learned directly from the data without any prior assumptions on physics and geological settings. The method is demonstrated on the Marmousi II elastic model, where a gas reservoir was successfully delineated from a background trend in a depth migrated image.

INTRODUCTION

In the most general sense, unsupervised learning is a subfield of machine learning that tries to infer hidden structure within an unlabeled dataset. Unsupervised methods are particularly useful when the inferred structure is lower dimensional than the original data. For example, given a list of n patients in a hospital and their corresponding symptoms s , it is unlikely that each patient-symptom combination is unique. A set of common diseases d can be inferred from the data, where $d \ll n, s$. Popular unsupervised learning and data mining methods such as principal component analysis (PCA) and K-Means clustering rely on exploiting low-dimensional structure inherent in the data (Ding and He, 2004).

Interestingly, interpreted images and geological maps produced by geoscience workflows are substantially lower dimension than the original field data. The structure of the major sedimentary layers of the Earth is relatively simple, as rocks with similar physical properties are formed along relatively continuous interfaces and facies in the subsurface. For this reason, we can use a combination of physical models, local geological knowledge, and experience to reduce large seismic and well-log datasets into low-dimensional models of the Earth. Abstractly, we are inferring a low-dimensional Earth model from high-dimensional geophysical data. In this respect, reservoir characterization can be posed as an unsupervised machine-learning problem.

In conventional AVA interpretation, two-term AVA attributes are extracted from seismic angle gathers using the Shuey approximation (Shuey, 1985) as a physical reflectivity model. Multivariate analysis of these attributes lead to an estimation of a background trend of shale-sand reflections and anomalous outliers that can be considered potential hydrocarbon indicators (Castagna et al., 1985, 1998). Although this has proven to be an effective workflow, the efficacy of the method requires calibrated seismic data processing that preserves reflection amplitudes throughout migration. In theory, amplitude preserving migration is feasible (Sava et al., 2001; Zhang et al., 2014; Gajewski et al., 2002), however there are always large uncertainty and variations in measured AVA responses. Recent work

by Hami-Eddine et al. (2012) applied neural networks to classify AVA anomalies, while Hagen (1982), Saleh et al. (2000), and Scheevel et al. (2001) used principal component analysis (PCA) to characterize pre-stack seismic data. We follow a similar philosophy and demonstrate that conventional AVA characterization can be reformulated as an unsupervised learning problem. In the vernacular of machine learning, the problem generalizes as dimensionality reduction followed by clustering.

THEORY & METHOD

Starting with angle-domain common-image gathers, we desire a segmented output image where each pixel is classified according to the local AVA response. We define the angle gathers as feature vectors $x_i \in \mathbb{R}^d, i \in [1, \dots, n]$, where n is the number of samples in the image and d is the number of angles in the gather. The feature vectors are shaped into a matrix $X \in \mathbb{R}^{n \times d}$, where each row corresponds to a point in the image and each column corresponds to an angle. Generalizing the data as a feature matrix allows us to work in an unsupervised learning framework (Figure 1).

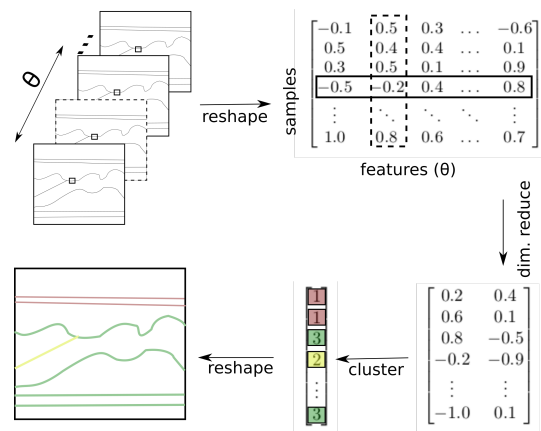


Figure 1: AVA characterization as unsupervised learning.

We assume that the columns of X are not independent, as the angle response of a reflection is often modeled by simple equations with as few as two parameters (e.g., two-term Shuey equation). Assuming the existence of a lower-dimensional representation, we can use dimensionality reduction techniques to reduce the number of columns in X into a new feature matrix $\hat{X} \in \mathbb{R}^{n \times m}, m \ll d$.

PCA reduces dimensionality by keeping the m most significant

eigenvectors from the decomposition of the covariance matrix

$$G = X^T X = \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \dots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_2, x_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \dots & \langle x_n, x_n \rangle \end{pmatrix}. \quad (1)$$

Although PCA will reduce the number of features while maximizing the variance of the data (a measure of information), it is a linear model which may not result in the best low-dimensional representation of X . Note that the covariance matrix G depends only on the inner product of the feature vectors $\langle x_i, x_j \rangle$ and not the features directly. We can thus replace the inner-products with a kernel function, which implicitly calculates a similarity measurement in a higher-dimensional feature space (Hofmann et al., 2008). Using a non-linear kernel function $\kappa(x_i, x_j)$ will result in a non-linear PCA operation (Schölkopf et al., 1997). In this study, we found by trial that the polynomial kernel

$$\kappa(x_i, x_j) = (x_i^T x_j + c)^d \quad (2)$$

with $c = 0$ and $d = 10$ provided the best clustering in our examples.

Assuming that common similarities in the rows of \hat{X} can be sorted into a finite set of groups, we can use a clustering algorithm to associate each sample to a group. Since there is no guarantee that the clusters will have Gaussian structure, methods that rely on Gaussian mixtures such as K-means are not appropriate for this application. Instead, we use BIRCH clustering (Zhang et al., 1996), which is a hierarchical clustering algorithm designed for large databases and makes no assumptions about underlying statistical distributions or cluster geometry. The output vector consists of the cluster identification number for each point, which is reshaped back into model dimensions resulting in a segmented image. Open-source software libraries Madagascar and scikit-learn were used for seismic processing and machine learning. All scripts are publically available at <https://github.com/ben-bougher/thesis>.

EXAMPLE

We tested the method using a subset of the elastic Marmousi II model (Martin et al., 2002) (Figure 2). This section of the model contains a gas reservoir embedded in layers of brine saturated sand and shales. We used the Zoeppritz equations to generate images of the true reflectivity response and also synthesized seismic gathers, which intentionally violate the amplitude preserving assumptions implied by conventional AVA analysis. The synthetic seismic was generated using visco-acoustic modeling and migrated using the sinking survey algorithm described by Sava and Fomel (2003), which does not preserve amplitude.

We ran the algorithm using PCA, kernelized PCA, and conventional two-term Shuey coefficients to reduce the datasets to two features. For the physically consistent data, all methods were able to cluster the background trend and anomalies, however the multivariate distribution of the reduced features showed interesting geometries. The Shuey terms (Figure 3)

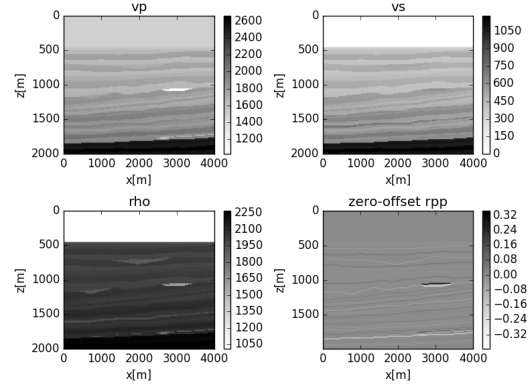


Figure 2: Subset of the Marmousi II elastic model.

and linear PCA (Figure 4) showed remarkably similar reduced features, where the kernelized PCA (Figure 5) yielded tighter more distinct clusters.

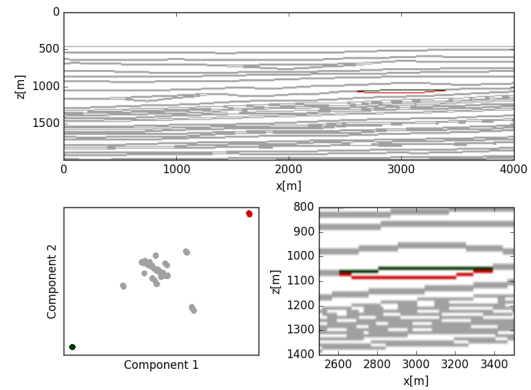


Figure 3: Clustering the true reflectivity model using Shuey terms.

The migrated seismic was peak filtered and thresholded to filter for reflection events. Shuey terms were extracted from the migrated seismic data using a basic least-squares data fit. The poor correlation between the Shuey coefficients reflect the physical inconsistencies between the model and the migrated gathers. Clustering on the Shuey terms was not able to discriminate the reservoir from the background trend (Figure 6); however, both the PCA (Figure 7) and kernelized PCA (Figure 8) showed significant delineation of the reservoir.

CONCLUSIONS

AVA characterization was presented in an unsupervised machine learning framework. PCA and non-linear kernel PCA feature reduction algorithms were compared to conventional Shuey coefficients. Each approach was able to segment the true reflectivity image, however the conventional Shuey term approach failed to delineate gas reservoir in the migrated seismic image. The main result of this work is that AVA analysis can

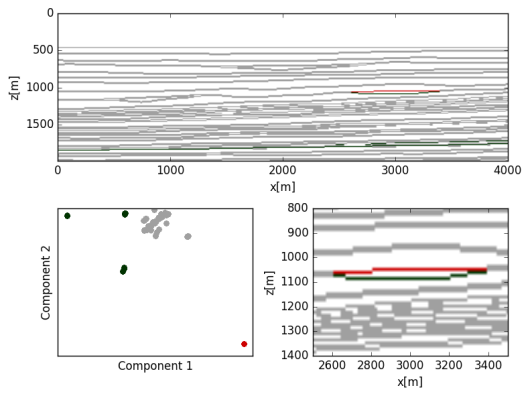


Figure 4: Clustering the true reflectivity model using principal components.

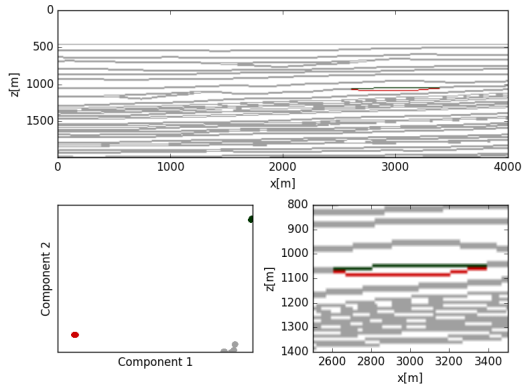


Figure 5: Clustering the true reflectivity model using kernelized principal components.

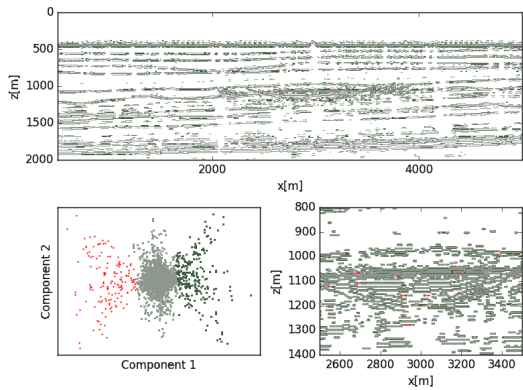


Figure 6: Clustering the migrated seismic using Shuey terms as features.

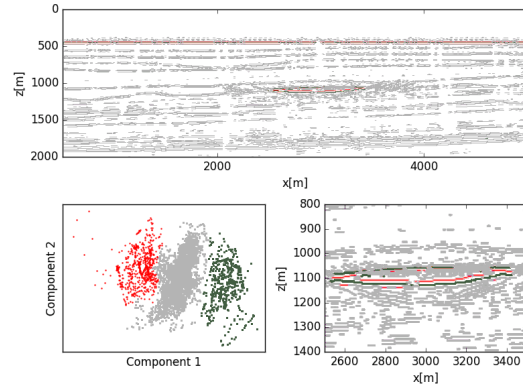


Figure 7: Clustering the migrated seismic using principal components as features.

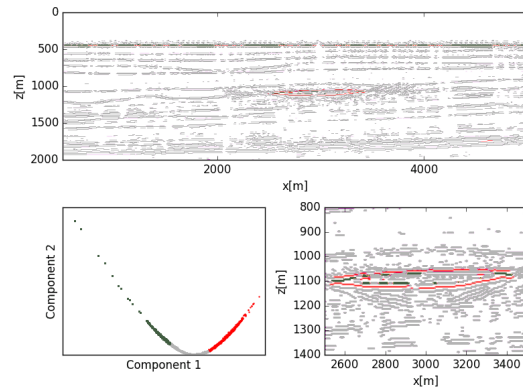


Figure 8: Clustering the migrated seismic using kernelized principal components as features.

be reformulated as a machine learning problem, which can successfully characterize an image without physical or geological assumptions.

ACKNOWLEDGEMENTS

This work was financially supported in part by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of the SINBAD II project with the support of the member organizations of the SINBAD Consortium.

REFERENCES

- Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985, Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks: *Geophysics*, **50**, 571–581.
- Castagna, J. P., H. W. Swan, and D. J. Foster, 1998, Framework for AVO gradient and intercept interpretation: *Geophysics*, **63**, 948–956.
- Ding, C., and X. He, 2004, K-means clustering via principal component analysis: Proceedings of the twenty-first international conference on Machine learning, ACM, 29.
- Gajewski, D., R. Coman, and C. Vanelle, 2002, Amplitude preserving Kirchhoff migration: a travelttime based strategy: *Studia geophysica et geodaetica*, **46**, 193–211.
- Hagen, D. C., 1982, The application of principal components analysis to seismic data sets: *Geoexploration*, **20**, 93–111.
- Hami-Eddine, K., P. Klein, L. Richard, and A. Furniss, 2012, Anomaly Detection Using Dynamic Neural Networks, Classification of Prestack Data: Presented at the SEG-2012-1222, Society of Exploration Geophysicists.
- Hofmann, T., B. Schölkopf, and A. J. Smola, 2008, Kernel methods in machine learning: *The Annals of Statistics*, **36**, 1171–1220.
- Martin, G. S., S. Larsen, K. Marfurt, and others, 2002, Marmousi-2: an updated model for the investigation of AVO in structurally complex areas: Presented at the 2002 SEG Annual Meeting, Society of Exploration Geophysicists.
- Saleh, S. J., J. A. de Bruin, and others, 2000, AVO attribute extraction via principal component analysis: Presented at the 2000 SEG Annual Meeting, Society of Exploration Geophysicists.
- Sava, P., B. Biondi, S. Fomel, and others, 2001, Amplitude-preserved common image gathers by wave-equation migration: 71st Ann. Internat. Mtg. Soc. of Expl. Geophys., 296–299.
- Sava, P., and S. Fomel, 2003, Angle domain common image gathers by wavefield extrapolation: *GEOPHYSICS*, **68**, 1065–1074.
- Scheevel, J. R., K. Payrazyan, and others, 2001, Principal component analysis applied to 3d seismic data for reservoir property estimation: *SPE Reservoir Evaluation & Engineering*, **4**, 64–72.
- Schölkopf, B., A. Smola, and K.-R. Müller, 1997, Kernel principal component analysis, *in* *Artificial Neural Networks—ICANN'97*: Springer, 583–588.
- Shuey, R. T., 1985, A simplification of the Zoeppritz equations: *Geophysics*, **50**, 609–614.
- Zhang, T., R. Ramakrishnan, and M. Livny, 1996, BIRCH: an efficient data clustering method for very large databases: *ACM Sigmod Record*, ACM, 103–114.
- Zhang, Y., A. Ratcliffe, G. Roberts, and L. Duan, 2014, Amplitude-preserving reverse time migration: From reflectivity to velocity and impedance inversion: *Geophysics*, **79**, S271–S283.