Machine learning in the interpreter's toolbox: unsupervised, supervised, and deep learning applications

Laura Bandura, Adam D. Halpert, and Zhao Zhang*
Chevron Energy Technology Company

Summary

Recent advances in both hardware and software have brought machine learning techniques to the forefront of many industries, disrupting traditional workflows and creating new opportunities. When applied to geophysical data and images, these same technologies have the potential to transform tasks ranging from data acquisition to processing, imaging, and interpretation. Here, we explore the latter category by presenting three applications of machine learning for seismic interpretation purposes: unsupervised clustering of image data according to AVA response; supervised classification of lithology or geomorphology based on seismic texture attributes; and seismic image resolution enhancement enabled by a "deep" learning algorithm. A common theme through each of these examples is the necessity for tight integration between the algorithms themselves and the geoscientific expertise required both to train the algorithms and to validate and interpret their results. As machine learning techniques become central to an increasing number of geophysical workflows, an emerging model of interaction between subject-matter experts and algorithms is one of a virtuous feedback loop – SME's provide their expertise to improve the performance of algorithms, which in turn can provide richer or higher-quality information than would otherwise be available to the SME's.

Introduction

Machine learning can aid in interpretation through the identification and characterization of underlying patterns in seismic and log data that are beyond human comprehension or obscured through tradition means of visualizing and interacting with the data. Machine learning tools can assist interpreters in making connections and correlations within and across datasets of different vintages, domains, and scales. As the complexity of the problems being addressed increases, the machine learning "hammer" in the interpreter's toolbox tends toward deep neural networks (DNN's), with network architectures that can be highly customized to return an optimal result.

Here, we highlight three machine learning techniques to identify, characterize, and visually enhance subsurface features (fluids, stratigraphic, structural, petrophysical) that are measurable in seismic, well log, or other geoscience data. These approaches have twin aims:

1) Automate tedious processes with a goal of efficiency, enabling interpreters to better utilize their time for developing a full subsurface story integrating all available data and knowledge;
2) Provide interpreters with richer, higher-quality data than would otherwise be available, allowing them to make better-informed decisions.

Full integration of subject-matter expert (SME) domain knowledge with these powerful analytical tools will enable optimization of individual and collective algorithms and workflows that are capable of learning new feature associations through time.

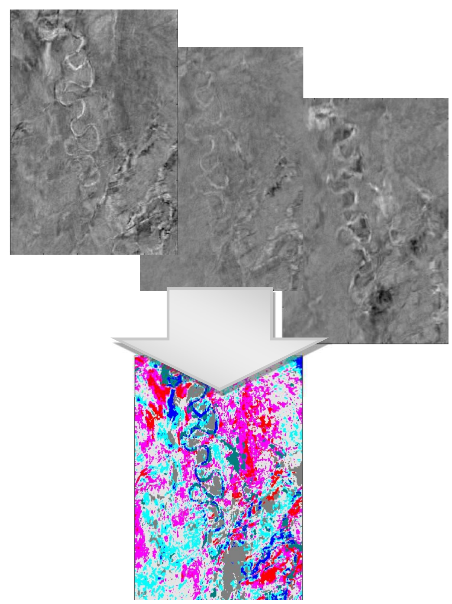

Figure 1: Pre-stack and partial stack seismic data can be segmented quantitatively with the aid of machine learning. Labelled cluster volumes derived from the pre-stack AVA response can be co-rendered with the stacked seismic data for enhanced, quantitative interpretation.

Unsupervised Learning

Unlike supervised machine learning methods, unsupervised techniques do not require the algorithm to be trained and, hence, no training data selection is needed. An unsupervised approach to seismic analysis is generally more straightforward and the evaluation process can rapidly yield direct data insights.

Unsupervised machine learning techniques such as clustering (Bezdec, 1981) or self-organizing maps (Kohonen, 2001) can aid in the visualization and, hence, the interpretation of multidimensional data. For example, standard amplitude vs. angle/offset (AVA/AVO) analysis is typically performed on a few partial angle/offset stacks. The geoscientist will typically view the seismic stacks side-by-side in an interpretation platform. Machine learning provides an opportunity to visualize the AVA/AVO response computed from multiple partial stacks or the full pre-stack gathers in one 3D volume (Figure 1).

Techniques such as clustering pre-stack seismic data can not only yield outputs that provide enhanced visualization, but they can also perform quantitative analyses on a large scale. These additional outputs can help interpreters assign deeper meaning to images and seismic volumes, conveying information such as fluid content or rock properties in addition to geomorphology (Figure 2) and geological context.

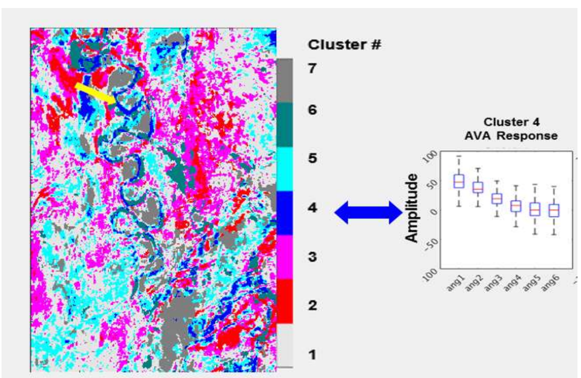

Figure 2: Horizon through a volume produced through clustering indicating the variety of AVA responses in the pre-stack seismic data. The global (volumetric) AVA statistics are displayed for cluster 4.

Supervised Learning

Seismic interpretation requires considerable experience and SME knowledge and must be done by skillful interpreters with intensive analysis. Supervised learning offers an opportunity to develop an expert-knowledge-based automated system, which incorporates both domain knowledge and quantitative data mining.

To achieve quantitative interpretation, supervised learning allows us to develop a systematic approach to identify and label stratigraphic components by describing them within a hierarchical framework that is based solely on the physical attributes of the strata and imaging attributes from the seismic image. For example, an interpreter can provide a machine learning algorithm with partial interpretations, either on the fly or through referencing a training data repository of prior interpretations. The algorithm extracts various patterns useful in interpretation from interpreter's input and applies the "learned" domain knowledge to a new dataset. The output identifies the most relevant information to present to the interpreter to guide next steps or decisions.

Seismic attributes can be important qualitative and quantitative predictors of reservoir properties and geometries when correctly used in reservoir characterization studies. In machine learning terms, these are the "features" used as inputs to an algorithm, and great care must be taken to select an appropriate feature set. For seismic image-based machine learning analyses, we advocate the use of textural image attributes, which describe the spatial arrangements of seismic segments and simulate the process experts use to interpret the seismic features. When paired with control data such as stratigraphic interpretations or physical reservoir properties measured at wells, these attributes can serve as reliable training data for supervised learning algorithms. By integrating seismic attributes, well log data, texture attributes and machine learning, supervised learning provides a unique opportunity to efficiently explore massive seismic datasets and uncover important insights.

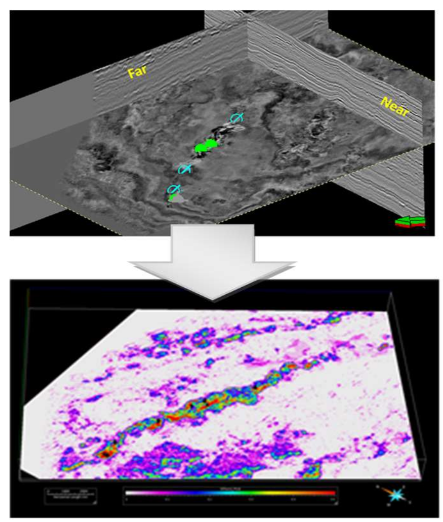

Figure 3: Guided by lithofacies logs, a machine learning based interpretation workflow provides oil sand distribution predictions throughout a West Africa seismic volume.

Figure 3 shows an example of a machine learning algorithm known as Random Forest (Breiman, 2001) applied to a West Africa dataset for the purpose of lithology and fluid identification. Here, training data consisted of seismic textural attributes labeled along well paths by a lithofacies log. The algorithm produces pixel-by-pixel probability estimates for each lithofacies type throughout the volume, allowing for straightforward, quantitative extraction of geobodies like the ones shown in Figure 4. This greatly facilitates further tasks such as inversion, earth modeling, and downstream property estimation QC processes.

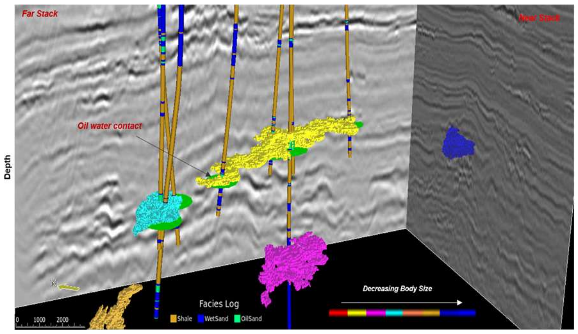

Figure 4: 3D view of extracted oil sand geobodies from the predicted oil-sand probability volume.

Deep Learning

Recent breakthroughs in both computational capability (for example, the use of general-purpose graphical processing units for scientific computation) and algorithmic theory have enabled a machine learning sub-discipline known as *deep learning* (LeCun et al, 2015; Goodfellow et al., 2016) to flourish. Deep learning approaches are characterized by sophisticated neural network architectures with many "hidden" computational layers that enable the networks to learn and apply highly nonlinear relationships between input and output variables.

Deep learning approaches have proven adept at image processing tasks. In particular, a deep network variant known as a *generative adversarial network* (GAN) (Goodfellow et al., 2014) has become quite popular for generating realistic, high-quality images when given a noisy or low-quality input. GAN's have great potential to improve the resolution and interpretability of seismic images, whose quality can suffer from a variety of issues, including:

- Physical limitations on resolution
- Limitations imposed by the computational expense of high-frequency imaging
- Processing and imaging challenges including the presence of coherent and incoherent noise, or illumination and velocity modeling issues.

Deep learning approaches such as a GAN offer an opportunity to partially overcome these obstacles in a completely data-driven manner. When provided with training data consisting of low- and high-quality image pairs from the same location in a seismic image, the network can attempt to infer relationships between patterns or textures present on a low-quality or distorted image, and the actual structures visible on its high-quality counterpart. A trained network can then be applied to previously-unseen data, allowing SME's to obtain greater insight from higher-quality, more interpretable realizations of seismic images than would otherwise be available.

To illustrate the potential of GAN's for seismic image enhancement, we apply a GAN based on the implementation of Garcia (2016) to an image obtained from the SEAM Phase I synthetic model (Fehler and Keliher, 2011). The GAN was trained with over 150,000 pairs of 128 x 128 pixel 2D patches extracted from the shaded portion of the image in Figure 5. Low-quality versions of the image patches were obtained by application of an aggressive low-pass filter to the original high-quality image. The trained GAN model was then applied to patches from a previously-unseen cross-section extracted from the un-shaded portion in Figure 5. After reconstructing the enhanced image patches, we obtain the result seen in Figure 6. While there are considerable opportunities for improvement, the GAN-enhanced result in panel b shows much closer resemblance to the ground truth (panel c) than does the low-resolution input image (panel a).

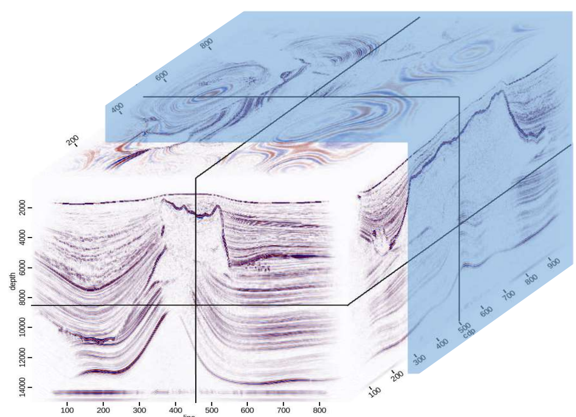

Figure 5: Training and validation data locations from the SEAM Phase I synthetic dataset. Training images were extracted from the shaded portion of the volume, while the test cross-section shown in Figure 6 was extracted from the unshaded portion, and not including in training.

Discussion and Conclusions

While machine (and particularly deep) learning will have significant business impact, there are also great risks. Critical challenges will include vetting the quality and reliability of the outputs of such efforts and the DNN's that create connections across functions and data. These systems will require active monitoring for accuracy as the scale and complexity of tools in the machine learning toolbox increases.

10.1190/segam2018-2997015.1

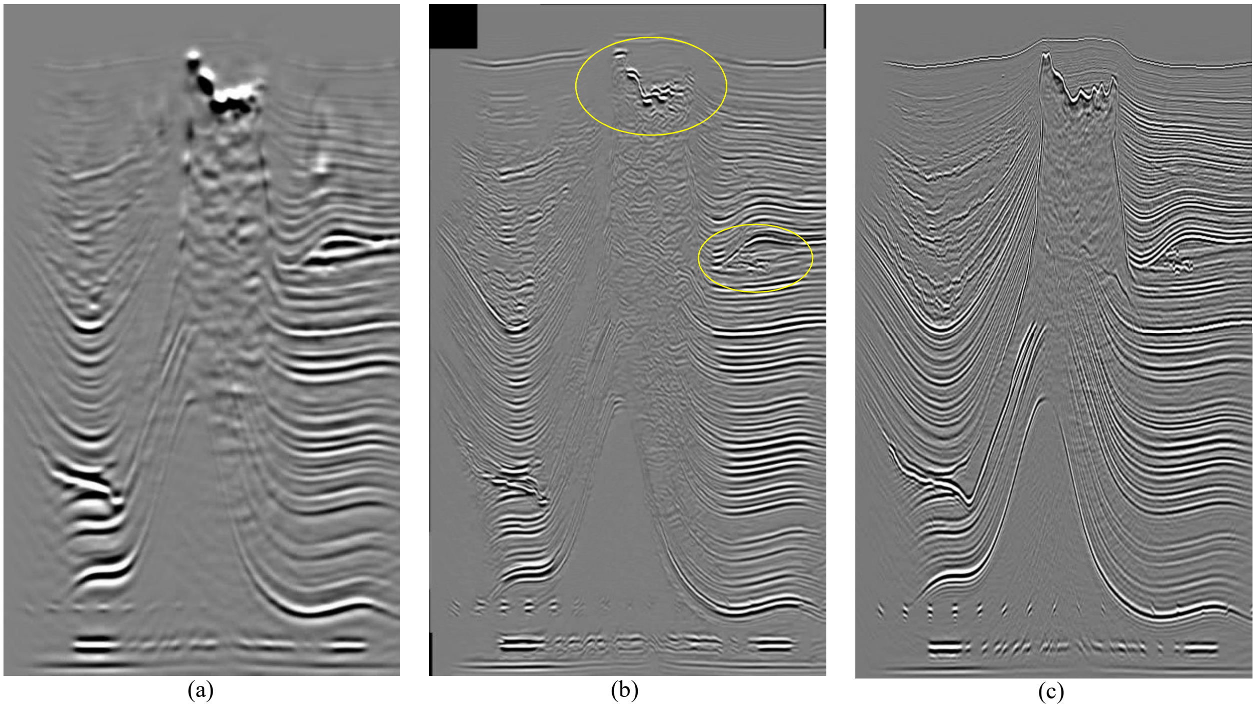

Figure 6: 2D example of image enhancement with a GAN. After patches from the low-resolution test image (a) are processed by the trained GAN, the reconstructed result (b) shows significant improvement in reflector sharpness and interpretability, especially in the indicated areas along the top salt boundary and a subsurface channel complex. Overall, the GAN result is more similar in character to the ground truth image (c) than to the input image.

There will be new roles created for earth scientists as a result of technology-driven transformations of many common workflows. These "hybrid" scientists will have talents obtained through cross-training in multiple earth science disciplines (e.g. geophysics, geology, petrophysics) and the most advanced machine learning and data science techniques. This new breed of earth scientists would be responsible for:

- Monitoring the neural network outputs for new discoveries of patterns/relationships revealed through the DNN (e.g. are the results physical?)
- Translating/conveying these discoveries from machine (computer) language to the earth science language of the SME's so that the value of such outputs is realized
- Working with SME's to introduce checkpoints and science-based constraints into the system
- Ensuring the evergreen DNN is up-to-date with any new knowledge, new data, and fit-for-purpose algorithms.

Our vision for successful deployment of machine learning technology is one where machine learning enables the derivation of maximum scientific insight and understanding from a system that connects data, concepts and people across multiple scientific disciplines for a level of impact that humans alone cannot deliver. We should not fear the machines, but rather embrace the exciting future where we can utilize their power for a greater understanding of the subsurface.

Acknowledgements

The authors would like to acknowledge Chevron Energy Technology Company and business units for permission to publish this work; the SEG Advanced Modeling (SEAM) consortium for creation of the synthetic model used in some of the examples presented here; and the following colleagues and collaborators for their advice and encouragement: Dimitri Bevc, Konstantin Osypov, Bruce Power, Tom Van Wagoner, David Sibley, Reynaldo Cardona and Ke Wang of Chevron, and Kenton Prindle, Eli Bixby, and Massimo Mascaro of Google. The authors are grateful to David Wilkinson for his inspiration on applying machine learning to the earth sciences.

REFERENCES

Bezdec, J. C., 1981, Pattern recognition with fuzzy objective function algorithms: Plenum Press.
Breiman, L., 2001, Random forests: Machine Learning, **45**, 5–32.
Fehler, M., and J. Keliher, 2011, SEAM Phase 1: Challenges of subsalt imaging in tertiary basins, with emphasis on deepwater Gulf of Mexico: SEG, e-book.
Garcia, D., 2016, Srez: Image super-resolution through deep learning, https://github.com/david-gpu/srez.
Goodfellow, I., Y. Bengio, and A. Courville, 2016, Deep learning: MIT Press 1.
Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, 2014, Generative adversarial nets, Advances in neural information processing systems, 2672–2680.
Kohonen, T., 2001, Self-organizing maps: Springer.
LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436.