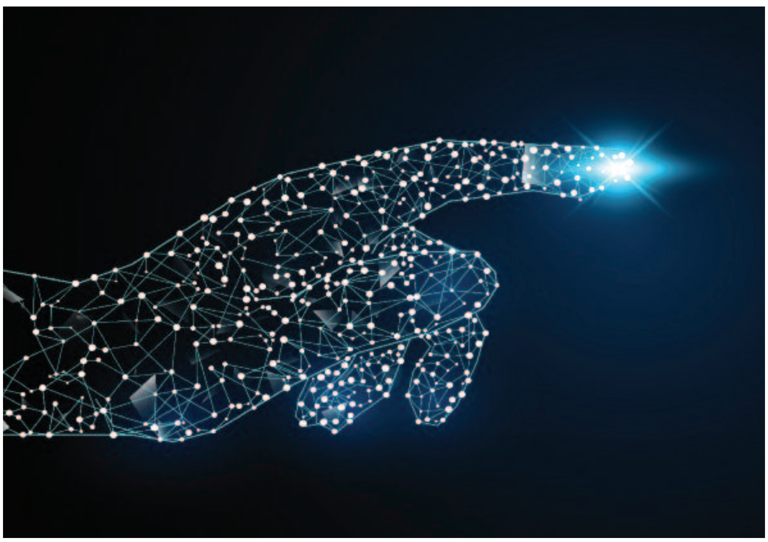
THE AMERICAN OIL & GAS REPORTER®
JANUARY 2018
The "Better Business" Publication Serving the Exploration / Drilling / Production Industry

Discovering New Horizons: Machine Learning Offers Faster, Better Interpretation

By Kamal Hami-Eddine,
Bruno de Ribet
and Alice Chanvin-Laaouissi

HOUSTON–Oil and gas exploration based on direct evidence of oil lasted for only a few years–at the very beginning of the industry's history–when oil fields were identified at locations with observable oil seepages. It did not take long for geologists to understand that bigger oil fields were located deeper underground, and were undetectable from the surface. Thus began the quest for indirect indications of hydrocarbon presence and traps.



The development of geophysical techniques has provided geologists with many tools to observe the subsurface in many resolutions. For example, logging tools provide local, high-resolution information about the subsurface, while seismic data enable geoscientists to obtain a lower-resolution–but global–view of the subsurface. We now have access to a multitude of direct and indirect measurements of the subsurface–more than a single human being can process and consume simultaneously.

While using statistical tools to help extract information from vast amounts of data is not new, the growth of the Internet and social media data have helped this methodology surpass all previous expectations. "Big data" analytical tools enable us to handle an unprecedented amount of information at unprecedented speed. Our industry will benefit from this in many ways.

Statistical approaches can be integrated into the subsurface evaluation process to analyze data and enhance collaboration between geoscientists. By doing so, we expand our knowledge to allow for reliable exploration decisions while minimizing risk. The rapidly increasing access to more real-time data will result in more information, thanks to improved machine learning techniques.

Adding Value

For many generations, people, or more specifically, state and religious representatives, have sought to register and measure information about land and people. We have testimonies regarding registries, tables and other types of summaries using typical values to characterize many types of phenomena.

The first known scientific application of statistical tools was actually in the field of earth sciences. As far back as 2,500 years ago, Babylonian astronomers were using compromised values to evaluate star positions, although the calculations behind them are unclear today. Later, in the second century, Greek astronomers such as Ptolemy deduced values from several measurements. They did not use the same statistics systematically, and the statistical approach was still very much an empirical process.

After centuries of statistical research and its application to social sciences, markets and earth sciences (among many other domains), we now have solid mathematical tools and experimental processes that enable us to analyze massive amounts of different types of data.

During subsurface evaluation, geoscientists attempt to identify the quality of a reservoir and its economic potential. To do so, they have access to large amounts of relatively high-quality seismic data and accurate well logging. Integrating seismic and well data always has been a challenge, and the ability to combine them, with their different responses and resolutions, provides better insight into the subsurface if it is done accurately.

Statistical tools, clustering techniques and neural networks provide different approaches to dealing with such data. For a geophysicist analyzing seismic data, unsupervised techniques may provide the information needed to make decisions about exploration or infill development well positioning. Combining structural and standard seismic attributes using principal component analysis (PCA) can provide better fault images, and thus a better understanding of potential trapping mechanisms.

Additionally, clustering different seismic attributes into seismic facies often provides stratigraphic patterns, which enables us to understand depositional systems. This leads to conclusions regarding rock type distribution, even before the well data are integrated.

Decoding Data

In some geological settings, amplitude-versus-offset/amplitude-versus-angle analysis can provide a good method for derisking target placement. An expected and observed AVA effect will increase the success rate of exploration drilling. Since AVA studies can be time consuming, automated AVO signature analysis can be performed using neural networks.

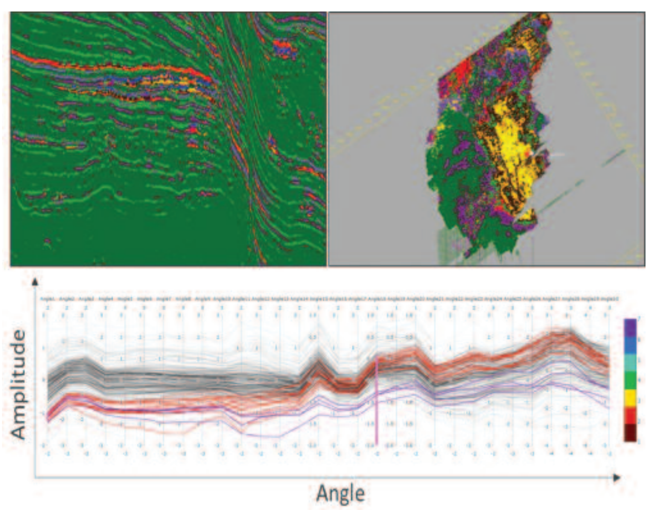
A 2017 paper by Kamal Hami-Eddine, Bruno de Ribet, Patrick Durand and Patxi Gascue finds that the use of self-growing neural gas enables fast-track identification of AVA responses through a neural network training and clustering process. This provides a way to map and locate potential AVA anomalies and regroup them for further analysis (Figure 1).

For this purpose, research by Bernd Fritzke published in 1995 shows an artificial neural network inspired by growing neural gas provides a good solution for matching high dimensionality in data such as prestack information, or for enabling the use of several post-stack seismic attributes.

However, like any self-organizing network, growing networks are attracted by data density. And considering that the goal is to overcome this issue for mapping anomalies, an analysis by Victoria Hodge and Jim Austin suggests the algorithm needs to be adapted to also focus on outliers. To reach this goal, the neural network training stage needs to:

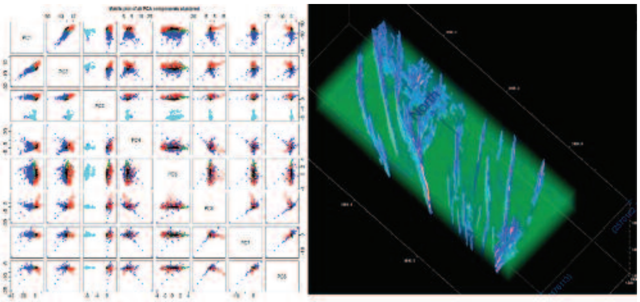
- Train the network on a sampled interval of interest;
- Detect outliers from a previous stage;
- Train only on outliers;

FIGURE 1



A clustering of time migrated angle gathers highlights different AVA responses in the Ariki formation. Each cluster is composed of different "modeled" AVA responses by the neural network as highlighted in the parallel coordinate plot.

FIGURE 2



This principal components scatter plot shows clusters isolating fault features in cyan. It is unnecessary to obtain all components to isolate faults.

- Cluster the neurons using a network topology approach; and
- Estimate the probability of belonging to a particular class.

These techniques provide prospect-specific tools that analyze data faster and transform massive amounts of it into understandable information for geoscientists. Techniques such as clustering and PCA (Figure 2) are ready for use, and already are well integrated into software platforms to analyze seismic data and well logs.

Tomorrow, though, may be different. What if neural networks become prospect agnostic? What if geoscientists obtain information directly derived from seismic data loaded into the interpretation platform they use daily?

Opening New Doors

In the last 10 years, machine learning has seen unprecedented development and a huge amount of investment. As anticipated in the early 1990s, the increase in data acquisition and the development of Web technology will bring new challenges. Hardware evolution in the 2000s has made it possible to train neural networks on enormous amounts of data, and these networks now surpass humans in some tasks, such as image recognition. Conventional neural networks could not have mastered such tasks easily, if at all. Deep learning methods have provided a new way to formulate problems and better imitate the way our brains work.

Deep learning's successes and failures show that it can add tremendous value to the oil and gas industry if we clearly understand it. Image recognition techniques are reaching new levels of accuracy thanks to convolutional neural networks (CNN), which enable on-the-fly analysis of terabytes of digital images acquired daily. This technology relies on three fundamental conditions:

- Fast data transfer and storage;
- Existing images with high-quality labels; and
- Trained deep networks available from anywhere in the world.

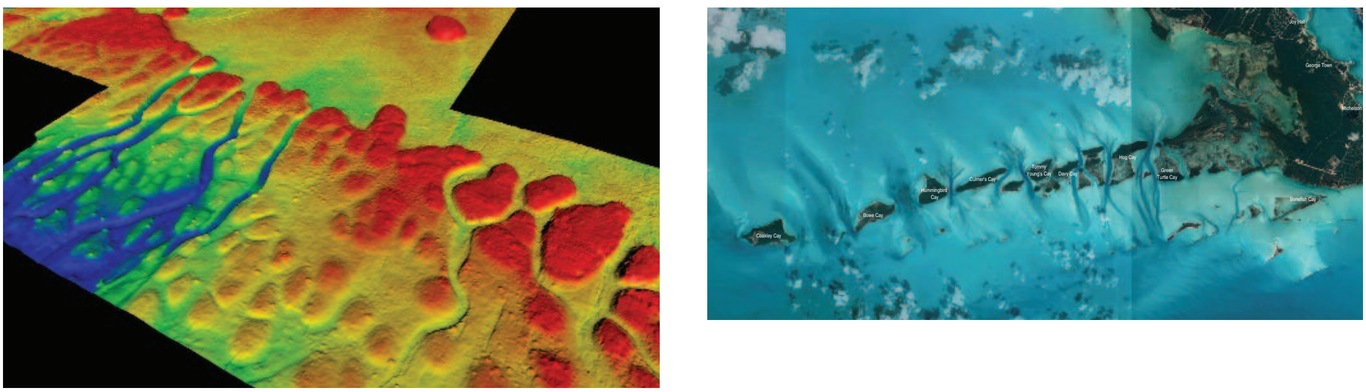
Seismic interpreters focus on labeling data. The processes of fault interpretation, mapping and geobody identification are labeling processes per se. The industry has volumes of labeled (previously interpreted) data that can be used to start training deep networks to automate tasks, including:

- Fracture characterization;
- Fault and horizon interpretation;
- Analog finding (Figure 3); and
- Document retrieval and analysis.

The principles of deep learning are relatively simple. Like any neural network, they rely on simple mathematical operations. Their strength lies in formulating the problem, making it possible to train multilayer networks from huge databases while limiting the risk of overtraining and vanishing gradients. This is achieved by following the old adage of "divide and conquer."

In the case of image detection, an image is split into subparts in the training's first stages so that each neuron does not see all

FIGURE 3



Many satellite images are available on the Web and can constitute a large database for surface analog identification using deep learning, such as an image of tidal channels cutting through Miocene reefs off Indonesia (left) and a satellite image from Bahama's reefs (right).

of the image's parts. Initial layers are not fully connected. At the same time, deep neural networks combine supervised learning with some unsupervised stages to identify image characteristics. In a sense, deep neural networks automatically calculate attributes from the original image before matching the labels and image locations.

Research by Lei Huang, Xishuang Dong and T. Edward Clee demonstrates that although fault interpretation can be time consuming, a deep neural network can reduce, if not completely avoid, the attribute calculation stages as the network defines and generates the intermediate attributes necessary to better discriminate fault patterns.

As deep learning systems are trained to be "context" agnostic, technology will evolve toward more data analysis right from the data loading stage. Seismic data will be processed and loaded into the interpretation system, giving geoscientists access not only to raw seismic data, but to fault probability attributes (and maybe other data) as well.

Combined with image segmentation techniques, the step toward proper automatic detection, which detects actual faults and not every single lineament, seems relatively small and reachable. These probability volumes will supply input to structural uncertainty modeling.

The algorithms in use today, of course, have limitations. They consume lots of memory, cannot be untrained, and require a lot of high-quality data. New network architectures that integrate new calculation units, such as astrocytes or capsule networks, provide interesting perspectives for the future.

The concept of astrocytes as new calculation units is interesting, and seems to improve network performance. Even though applying them to deep networks seems distant, the idea is seductive. Capsule networks, on the other hand, seem to offer a promising path for overcoming spatial hierarchy issues. This may lead to a network architecture that can be trained with much less data.

Still in their infancies, these approaches have performance issues, which are even a major problem for most of the common deep learning techniques such as CNNs. Bridging high performance computing and machine learning also will determine how far the industry can go in using deep learning in the coming years.

In the meantime, machine learning is useful, and the day deep learning will help our industry seems imminent. However, a few requirements must be met before it becomes a reality. As these systems learn from data, the data need to flow and show some consistency in content and format. Making sure databases are clean, standardized and accessible is critical to unlocking the power of deep learning.

FIGURE 4

Users, parameter tracking and automation engines need tight access and connection to data.

Domain Cross-Fertilization

Considerable geophysical expertise is lost every year and the field's population age pyramid is skewed, making it obvious that our industry needs to make intensive and optimal use of data. Being able to retrieve knowledge from data is critical; it is not only a matter of improving process speed, but also a way to enhance efficiency in geological and geophysical workflows.

Expertise in oil exploration relies mainly on a good understanding of the tight relationship between geophysics and geology acquired through time by seismic interpreters and reservoir modelers. The life cycle of the subsurface evaluation process is a long story of data exchange (Figure 4) and analysis transferred from one expert to another to match the geological hypothesis against acquired and interpreted data.

In that sense, prior to any automation, the data flow between domains and users is critical. Within the geoscientist community, different specialists need access to the information associated with any project, sometimes simultaneously. An expert's interpreted or labeled data are needed–unduplicated–by the next expert who works on the project, and by keeping track of data labeling, quality control and modification throughout the full oil field cycle–which is also known as data traceability.

This mindset leads to a preference for multiuser and open data storage based on open data standards so that data can be accessed and used by any specialist, and possibly, any machine learning-based process. Efficiency also can be improved by considering real-time modifications or data arrivals and updates at any stage. In the same way machine learning is able to perform image detection on videos at 60 frames a second, it should be able to geosteer an exploration well from real-time drilling data and a static reservoir model.

While there is a need to connect geophysical, petrophysical and geological data, unstructured documents such as reports also need to find a place in this new ecosystem. Data and metadata constitute the information on which geoscientists base their decisions. How efficient would it be to consider analysis reports about cuttings in addition to logs to create electrofacies logs? The impact of considering this information at the upscaling stage, in order to integrate facies in a reservoir model, may accelerate the process and lead to a more petrophysically compliant model.

This approach will add significant value for oil companies, but we can leverage even more value by sharing across multiple companies. More data, and more variability in that data, make for more efficient machine learning-based systems. Ultimately, companies will not have to share or make their own data public, but simply will provide access for training the machine. Once the machine is trained, no original data information can be recovered.

Conclusion

The question no longer centers on whether artificial intelligence and machine learning will benefit our industry. Rather, it is a matter of how and when.

Machine learning techniques are available today to accelerate interpretation processes and account for more data when working on a prospect. In many cases, they provide faster subsurface images while still maintaining accuracy to help improve the decision making process. Conventional machine learning techniques do not make decisions by themselves, but they are helpers.

In order to fully move toward using deep learning in oil and gas exploration, companies and geoscientists need to engage in discussions with technologists who specialize in software development, data management and AI research.

Defining and formulating problems to be solved with machine learning will require close collaboration between data scientists and subject matter experts–geophysicists, geologists and engineers–so that AI systems can learn from the implicit knowledge known as experience to solve basic tasks. This will lead to systems that enable geoscientists to concentrate on issues requiring a deeper geological understanding.

The gap between machine learning and proper AI, which could replace humans, remains huge. At this point, we are dealing with algorithms to simplify our lives. In order to reach the level of AI, these techniques will have to improve in aspects that include:

- Memory;
- Retraining without the need to start over;
- Amount of data required to be trained;
- Decision making; and
- Unsupervised training.

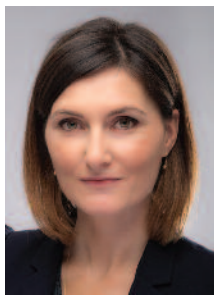
In the meantime, we can make good use of machine learning to accelerate processes. ❐

KAMAL HAMI-EDDINE

Kamal Hami-Eddine is product manager at Paradigm, now part of Emerson Automation Solutions. One of the inventors of rock type classification, he served as research scientist and technical consultant for the company prior to assuming his current role. Hami-Eddine holds a masters in applied mathematics and stochastic computations from the "Institut National des Sciences Appliquées" of Toulouse in France.

BRUNO DE RIBET

Bruno de Ribet is technology global director at Paradigm, now part of Emerson Automation Solutions. As a key subject matter expert, de Ribet advises on developing interpretation and reservoir characterization workflows that increase efficiency and take full advantage of the Paradigm suite of geophysical and geological innovations. He has more than two decades of experience in the upstream oil and gas industry and holds a Ph.D. in geophysics from the Institute du Physique du Globe, Paris University.



ALICE CHANVIN-LAAOUISSI

Alice Chanvin-Laaouissi is product manager at Paradigm, now a part of Emerson Automation Solutions. For the past three years, her focus has been on data management and data model integration for Paradigm solutions, which include the aim of integrating modeling data into relational database management systems. Chanvin-Laaouissi holds a master's in geology from the Ecole Nationale Supérieure de Géologie in Nancy, France, where she specialized in reservoir modeling in the oil industry.