Neural Networks Find Meaning In Data

By Brian Russell

CALGARY–The ultimate goal of any seismic analysis is to extend reservoir knowledge beyond the wells. Neural networks represent a class of statistical methods that yield rich detail that can greatly assist in reservoir characterization and planning. While nothing is ever certain until a well is drilled and completed, the chances of success are noticeably improved with this approach.

Neural networks are being used in just about every discipline where large volumes of data must be analyzed, including financial forecasting, medical imaging, and an amazingly wide array of other applications in science and technology. First conceived as a way to mimic how the human brain works, neural networks focus on minimizing the error in predictions and processes. These methods have moved out of the research lab and into commercial application, even as research into the original human neural pathway research continues. They have been shown to improve the accuracy of lithological predictions in seismic analysis and are an important tool for geoscientists.

Neural networks are well suited to geoscience challenges, given that they can solve for many variables at one time and can handle the very complex statistical analysis required to create a prediction of lithology from large seismic data volumes. There are a variety of different types of neural networks, but the one that is best known is the multilayer feed forward (MLF), or back-propagation, neural network. (Although these names sound contradictory, they come from the fact that the data is fed forward through the network, but the network training is done by back-propagating the error.)

The foundational element of the MLF network is the neuron, which takes the input, applies a nonlinear function to it, and then produces an output. Typically, this is a sigmoidal type (s-shaped) function that turns on or off in a slowly decaying way. This nonlinear concept seems straightforward today, but it took more than 40 years for researchers to develop the concept. The nonlinear nature of the neuron is what distinguishes this type of neural network from a multilinear regression. This nonlinearity can reveal hidden features in the data, but must be treated with care so as not to come up with false predictions.

For seismic analysis, there are two major applications for neural networks. The first is sample-to-sample prediction, where the goal is to identify every sample on a 3-D seismic volume with a specific lithological or fluid parameter such as porosity or water saturation. The input to this type of analysis can be a wide variety of seismic attributes, but we find that the best input attributes often consist of the output of a prestack seismic inversion, specifically the compressional (P) impedance, shear (S) impedance, and density. Sample-to-sample prediction is used when the geoscientist wants to see how key reservoir values change at every point across the data set.

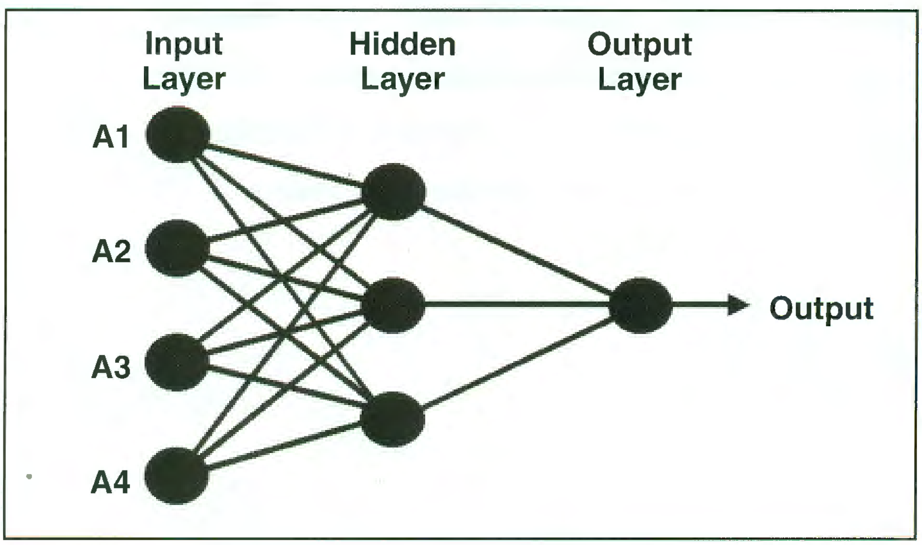
The second application is classification. In classification, the goal is to identify a series of classes or lithofacies. The geoscientist initially divides the zone of interest into classes such as sandstone, shale and carbonate, or water sand, gas sand and oil sand. The neural network then proceeds to identify these classes throughout the seismic data volume. Classification also is used when the focus is on groupings such as high, mid and low values for a particular attribute, say porosity. Classification also provides probability or uncertainty estimates that do not come with sample-to-sample prediction.

In both cases, the output of the analysis is a more geologically meaningful result than the original seismic inversion input. The level of trust that the model inspires depends on original data quality, ground truthing and cross-validation. In cross-validation, selected samples or wells are left out of the training procedure and the computed values are then compared with the known (and left out) values.

The schematic diagram in Figure 1 shows the multilayer feed forward network, where A1 through A4 are the input seismic attributes that are fed into the input layer. The circles in the hidden layer are nonlinear neurons, and the lines represent weight values. The output value is trained to match the desired reservoir parameter.

FIGURE 1

Multilayer Feed Forward Network



Unbiased Analysis

From a practical perspective, neural networks are a subgroup of advanced statistical methods used to find meaning in large volumes of data. The many types of neural networks referred to earlier are focused on different kinds of problems. The neural networks discussed in this article are supervised neural networks. By supervised, we mean that we train the network by presenting it with observed values (say, well porosity) at locations on the seismic volume that tie to the well and then let the neural network determine a relationship between the seismic attributes and the well values.

There are also unsupervised neural networks, which look for underlying patterns in the data but do not tell the interpreter specifically what these patterns mean. An example of such a neural network would be the Kohonen self-organizing map.

A chief distinction from deterministic seismic analysis methods such as inversion and AVO is that multivariate statistics, classification and neural networks do not assume an underlying model, but instead try to build the relationships directly from the data. This distinction is significant because it removes potential bias from the analysis.

For example, the geophysical model behind post-stack inversion is that the seismic trace represents a wavelet convolved with a reflection coefficient series that is derived from the P-impedance. Post-stack inversion, therefore, tries to recover the P-impedance. Prestack seismic data inversion extends this by using the Aki-Richards linearized model to interpret the data as a function of angle, P-impedance, S-impedance and density.

Even when interpreters are simply looking at a seismic section, they have some kind of "model" in their heads, such as, for example, when they follow a reflection between a sandstone and a carbonate. They may not be inverting the data in a mathematical way, but they automatically see boundaries between layers. No computer will ever replace a good interpreter (thank goodness!) and there are times when the "bias" that the interpreter introduces into his interpretation is vastly superior to anything a computer can do, based on the interpreter's years of experience in looking at data in a particular area. However, there are times when this bias can hurt the interpretation, and we need to rely on an "unbiased" computer.

One of my favorite exercises when I am teaching these techniques is to use them to predict a shear-wave sonic log from a weighted combination of other logs. I first ask the students to predict the order of correlation they would expect between the shear-wave log and the P-wave sonic, density, and gamma ray, and they predict that P-wave sonic and density should correlate well, but gamma ray should not. We then let the computer use multivariate regression to make the prediction and find that a better solution is arrived at when the density log is dropped out of the calculation, and only the P-wave sonic and gamma ray logs are used in the regression. That is, the best solution is not physically intuitive, and the program comes up with a better prediction because it is totally unbiased.

Figure 2 displays the results from a multilinear regression (top) and a neural network (bottom) to predict reservoir porosity from multiple seismic attributes, as shown on a seismic section. Note the higher resolution given by the neural network.

Figure 3 shows the results from a multilinear regression (top) and a neural network (bottom) to predict a low-velocity sand channel from multiple seismic attributes, as shown on a map derived from a seismic volume. Although the results are similar, the channel is more sharply defined by the neural network.

Sample-To-Sample prediction

Neural networks can increase both the level of detail and the degree of confidence the results engender. Sample-to-sample prediction makes a "best fit" prediction of reservoir properties at every seismic data sample. This is in contrast to stochastic methods, which produce multiple potential models that are equally probable. Another statistical approach is geostatistical or spatial analysis, which allows the geoscientist to build a map or volume of the predicted reservoir parameters between wells. Although not discussed in this article, stochastic and geostatistical methods are powerful tools that should be in every geoscientist's arsenal of techniques.

In sample-to-sample prediction, the first phase is training the multilinear regression or neural network at the wells, where the more wells we have, the better the result. A series of attributes are selected from which to derive a parameter of interest such as density, water saturation, porosity, or clay content. The attributes can be derived from the seismic volume itself, such as the amplitudes of the seismic traces or the instantaneous phase, instantaneous frequency, and derivatives of seismic trace. Attributes also can be inversion-derived such as P-impedance, S-impedance, and density, as mentioned earlier.

Even though 50 or more attributes could be used as inputs to the analysis, it can be shown that this will lead to "overtraining" (that is, false prediction), and so the cross-validation technique referred to earlier is used to find the statistically meaningful attributes, usually fewer than 10.

Once the attributes are selected, weights are assigned to each of them. In the multivariate regression approach, the weights are determined by a "least-squares" procedure, and can involve a convolutional operator to account for frequency differences between the well and seismic values. In the neural network approach, the initial weight setting is typically a random number. The output based on these initial settings is compared with known wells to determine errors, which are then fed back through the network to determine a better weight for each attribute. This iterative training process is designed to minimize errors and produce the best estimate of each attribute across all samples in the seismic volume.

As mentioned, most statistical techniques use a least-squared approach to minimize errors. Neural networks always

FIGURE 2

Predicted Reservoir Porosity from Multiple Seismic Attributes from Multilinear Regression (Top) and Neural Network (Bottom)

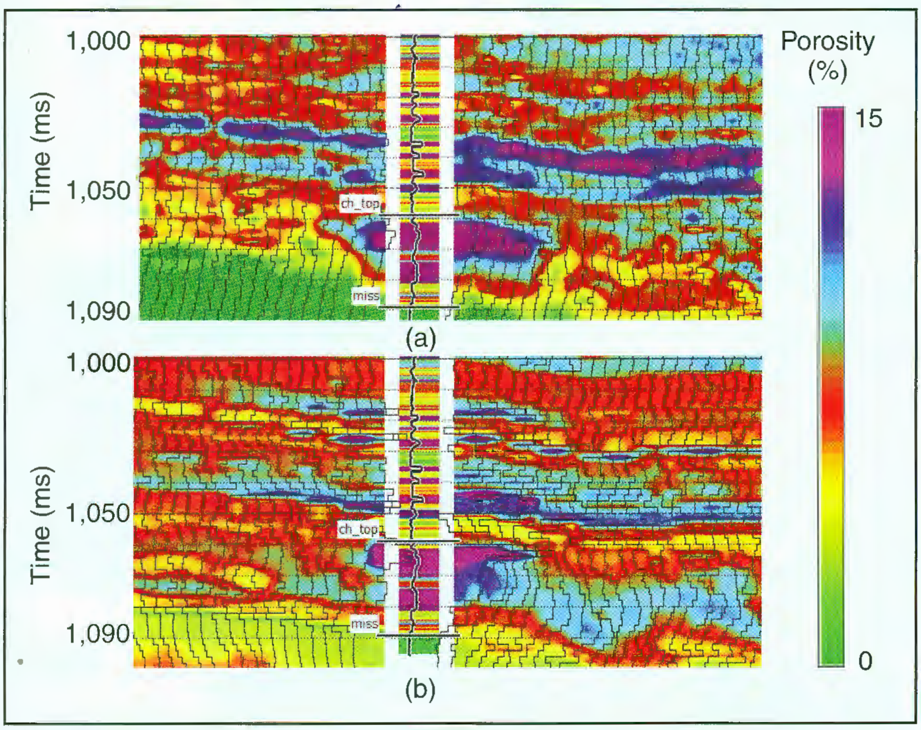
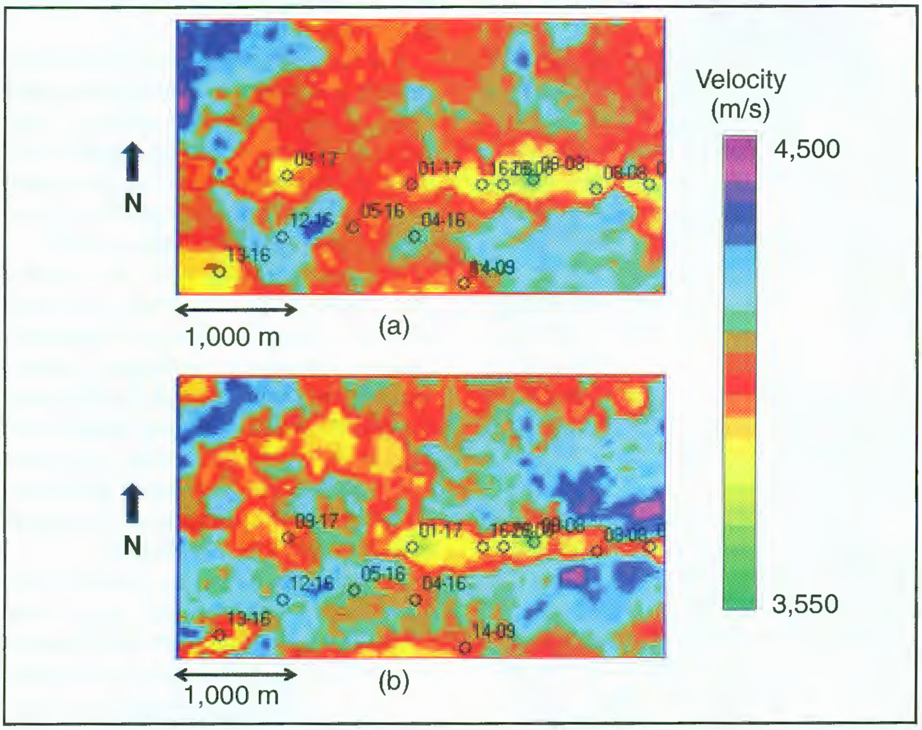


FIGURE 3

Predicted Low-Velocity Sand Channel from Multiple Seismic Attributes from Multilinear Regression (Top) and Neural Network (Bottom)



appear to do a better job, and the initial correlation is often extremely accurate. However, there is a danger of overtraining that is much worse than with linear statistical methods. Cross-validation is crucial in determining whether the initial correlation is representative or overdone, which can result in false predictions.

Classification

Classification is sometimes preferred to sample-to-sample prediction, because in addition to providing detail about specific lithofacies of interest, it also provides the probability of their presence in any given portion of the seismic data set. As a prerequisite, the asset team must determine what lithotypes they will seek, such as carbonates, shales, wet sands, oil sands and gas sands. These are quite common, but for any given analysis, there may be a few other facies of interest, so there could be five to seven total classes to define.

After the lithofacies are defined, the neural network must be trained to recognize them. This training concept stems from the early days of brain research, where scientists endeavored to replicate human learning. We learn from experience, and for geoscience problems, that experience comes from the wells. We believe the wells; they tell us more than the seismic. However, the well data are much sparser than the seismic data. So, in effect, the geoscientist trains the seismic to make the best possible fit to the wells.

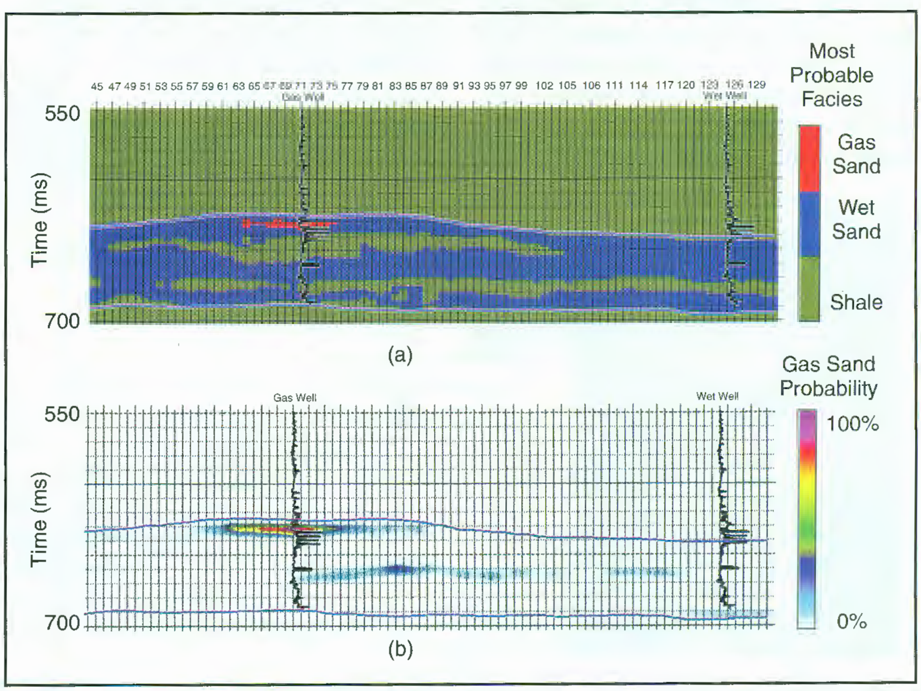
The training and prediction done in the sample-to-sample method gives the geoscientist a best-fit solution, but no-confidence estimates. As mentioned, classification differs in that it will give us a confidence, or probability, estimate. For example, in a particular area, there may be a 90 percent likelihood of finding a gas sand versus a wet sand.

The techniques used in classification involve either parametric or nonparametric statistics, or the probabilistic neural network, which turns out to be virtually identical to nonparametric statistics. In traditional parametric statistics, the parameters involve the mean and variance in one dimension, and the multivariate mean and covariance in higher dimensions. Also, the Gaussian probability distribution usually is used, which has the familiar "bell-shaped" look in one dimension, becomes elliptical in two dimensions, and is a hyper-ellipse in multidimensions.

In nonparametric statistics, we fit each

FIGURE 4

Seismo-Facies Clustering Showing Most Probable Facies (Top) and Probability of Finding Gas Sand (Bottom)



data point with a Gaussian function (or some other well-defined function) where the fit to the data is determined by a scaling parameter similar to the variance or covariance. We also can distinguish between Bayesian and maximum-likelihood classification, where the distinction is that Bayesian statistics add a prior probability function, which is usually derived from the ratios of the number of samples found in each class. By building in the statistics, we can obtain estimates of the probability of our prediction.

A good way to think about this is as a scatter pattern of observed values (say, porosity being compared with one or more seismic attributes). In one dimension, a straight line fit gives us a reasonable estimate of porosity (for example, P-impedance will increase as porosity goes down). In two dimensions, a flat planar surface again gives us a reasonable fit (for example, P-impedance and density versus porosity).

However, the line or plane will never fit the points perfectly. A bell-shaped curve (in one dimension) or an ellipse (in two dimensions), with their associated means and variances, does a much better job of fitting most of the data. Utilizing these statistical shapes to fit the scatter in the data is what allows us to predict their probability of occurrence.

Figure 4 is an example of seismo-facies clustering as applied to a seismic line showing (at top) the most probable facies for gas sand, wet sand and shale, and (at bottom) the probability of finding a gas sand. This analysis was done with Bayesian nonparametric statistical clustering, which, as mentioned earlier, is virtually identical to a probabilistic neural network.

Conclusion

As we have discussed, neural networks can be used to predict rock property behavior away from well control, which is a critical aspect in characterizing shale plays. Input data often include pore pressure prediction, stress analysis, curvature, the results of simultaneous seismic inversion and other seismic attributes. Training drives the optimum combination of attributes to predict volumetric petrophysical properties, including total porosity, effective porosity, clay volume, water saturation, quartz volume, and carbonate volume. Again, cross-validation using blind wells ensures that the network is not overtrained.

A study in the Haynesville Shale showed that no single attribute provided enough information to correlate seismic data to production. However, a predicted production map with 95 percent correlation was developed by correlating multiple attributes to average production and horizontal lateral length at different well locations. Drivers of the correlation were Young's modulus, differential horizontal stress ratio, Poisson's ratio and density.

The geoscientist has to know how far he can push the data. Challenging geological environments typically are the target for advanced analysis, but the analysis can be only as trustworthy as the measured data. If the geoscientist trusts the AVO data, then the attributes will be very good to take into the neural network. When working in a highly thrusted area, for example, he should be cautious about data integrity. A classic interpretation, where the geoscientist relies on years of interpretational knowledge, may be the best option in this case.

For the methods described in this article to work, everything depends on the quality of the seismic amplitudes. Analysis using low-quality data is what has given many of these techniques a bad name in the past. Applying high-quality analysis to low-quality data gives meaningless results.

On the flip side, with high-quality data, multivariate statistical and neural network techniques can deliver astounding clarity. Supervised techniques provide ground truthing from the well data, meaning that the results are more trustworthy. Output from classic seismic inversion methods, especially prestack inversion, provides a good starting point for further analysis using a neural network.

Both sample-to-sample prediction and classification can take the same input (for example, Vp/Vs ratios and P-impedance estimates), allowing the geoscientist to decide what approach output is most appropriate for each analysis.

As a final summary, these new techniques should never replace good interpretation practice. Rather, they should be seen as valuable companion tools when used effectively by the skilled interpreter. ❐

Coming In February
Special Report:
Unconventional Resources, Part II

BRIAN RUSSELL

Brian Russell is vice president of Hampson-Russell Software & Services, a wholly owned subsidiary of CGGVeritas. He entered the industry in 1976, and worked for Veritas for four years before co-founding Hampson-Russell Software & Services in 1987. Russell is actively involved in both geophysical research and training, and regularly presents courses on seismic technology throughout the world. A past president of both the Society of Exploration Geophysicists and the Canadian Society of Exploration Geophysicists, he has been awarded the SEG Enterprise Award, the CSEG Meritorious Service Award, and the CSEG Medal. Russell holds a B.Sc. in geophysics from the University of Saskatchewan, an M.Sc. in geophysics from the University of Durham, U.K., and a Ph.D. in geophysics at the University of Calgary.