Machine learning and learning from machines

Ehsan Zabihi Naeini[1] and Kenton Prindle[2]

https://doi.org/10.1190/tle37120886.1.

Abstract

Machine learning has been around for decades or, depending on your view, centuries. To consider the tools and underpinnings of machine learning, one would need to go back to the work of Bayes and Laplace, the derivation of least squares, and Markov chains, all of which form the basis and the probability construct used pervasively in machine learning. There has been a flood of progress between 1950 (with Alan Turing's proposal of a learning machine) and early 2000 (with practical applications of deep learning in place and more recent advances such as AlexNet in 2012). Deep learning has demonstrated tremendous success in a variety of application domains in the past few years, and with some new modalities of applications, it continues to open new opportunities. The recent popularity and emergence of machine learning in the oil and gas industry is likely due to the abundance of unused or overlooked data and the economic need to extract additional information from the data currently used. Additionally, there is an unprecedented availability of computing power, easy-to-use coding libraries, and application programming interfaces, as well as recent and significant advances in various flavors of neural networks. In this paper, we will attempt to show how machine learning can assist geoscientists in performing routine tasks in a much shorter time frame. We assert that there is a great opportunity for geoscientists to learn from machines, use these techniques to quality check their work, and gain nuanced insights from their data. Another advantage is that these approaches lead to the optimization of machine learning workflows by providing more accurate training data sets thus driving continuous learning and enhancement of the model.

Introduction

Machine learning has been around for decades, or depending on your view, centuries. If one were to consider the tools and underpinnings of machine learning, one would need to go back to the essay on Bayes's theorem (Bayes and Price, 1763; Laplace, 1812), which derives the probability construct used pervasively in machine learning, the derivation of least squares (Legendre, 1805), and Markov chains (Markov, 1906).

The first practical and somewhat visionary applications of machines and learning started in 1950 with Turing's proposal of a "learning machine" (Turing, 1950), the building of the first neural network machine known as SNARC by Marvin Minsky and Dean Edmonds (Crevier, 1993), and a machine capable of playing checkers that was built by Arthur Samuel at IBM (McCarthy and Feigenbaum, 1990). After the initial push in the early 1950s, a flood of progress was made toward modern machine learning including the invention of the perceptron (Rosenblatt, 1958); automatic differentiation and back propagation (Linnainmaa, 1970; Rumelhart et al., 1986); introduction of the concept of term frequency-inverse document frequency for natural language processing (Jones, 1973); neocognitron, which was a type of artificial neural network and the launch pad of convolutional neural networks (Fukushima, 1980); recurrent neural network (Hopfield, 1982); Q-learning, which effectively enables the concept of reinforcement learning (Watkins, 1989); random forest algorithm (Ho, 1995); support vector machines (Cortes and Vapnik, 1995); and long short-term memory (Hochreiter and Schmidhuber, 1997). Other notable achievements that have enabled the fast uptake and progress of machine learning include open source data sets such as MNIST (LeCun et al., 2018) and ImageNet (Deng et al., 2009). We mention them because we believe these efforts have led to the exponential growth in machine learning in the last decade (ideally, we would need such reference data sets for geoscience). As an example, AlexNet (Krizhevsky et al., 2012) was the winner of the ImageNet competition in 2012 where the authors invented a convolutional neural network structure that outperformed the runner-up by more than 10% accuracy. Since then, AlexNet has had a big impact on the field of machine learning, more specifically in the application of deep learning to computer vision.

Deep learning has demonstrated tremendous success in a variety of application domains in the past few years, and with some new modalities of applications it continues to open new opportunities. We see applications of machine learning in our daily lives, stretching from familiar applications such as spam filters dating back to the 1990s to more advanced cases such as self-driving cars and the automation of medical imaging and diagnoses. Behind any machine learning application could be any of the above-mentioned algorithms (e.g., naïve Bayes, decision trees, random forests, k-nearest neighbors, neural networks, etc.). However, one could argue that neural networks have undergone revolutionary progress and that advances in deep learning and success cases of convolutional and various flavors of recurrent neural networks have led to today's hype surrounding machine learning implementation. It is not the intention of the authors to explain all these algorithms in detail here. Instead, the focus is on demonstrating an application of machine learning and the possibility of learning from machines in typical geoscience applications. For example, we will demonstrate learnings from use cases such as document and image segmentation, facies classification, petrophysical log prediction, and fault interpretation. These applications are essential parts of various geoscience workflows, and although not too complex in theory, the machine learning approaches can certainly save time (and therefore money).

While there is a lot of enthusiasm in implementing machine learning for various geoscience applications, there is also skepticism

[1]Ikon Science Ltd., London, UK. E-mail: enaeini@ikonscience.com.
[2]Google Inc., Austin, Texas, USA. E-mail: kentonprindle@google.com.

of "black box" applications and how to quality control (QC) the outcomes. There are various QC measures in data science that are implemented routinely. One robust approach is cross validation, which is mostly known by geoscientists as blind well QC for well-based applications. It is discussed here that such approaches can lead to the concept of "machine learning and learning from machines." This concept is shown in the use cases mentioned earlier. The examples shown are selected so that the concept behind this paper can be demonstrated. We finish the paper with discussions around the computational complexities involved in machine learning applications in geoscience.

Document and image segmentation

We will begin with an example from an image segmentation and classification algorithm in which the method segments all the text, tables, and images in all the uploaded documents. The application of such tools to find keywords and relevant figures among hundreds and sometimes thousands of documents is phenomenal in saving time for practitioners. Some of the outcomes of this algorithm on two papers are shown in Figure 1. When applied to a paper by Infante-Paez et al. (2017), the algorithm correctly segments various figures. The subsequent classification outcome is somewhat amusing as it classifies the human photo as "thin micrograph limestone section" (Figure 1a). This is simply due to the fact that there was not a label class for human photos during training. Hence, the algorithm finds the most similar color scale and segment distribution, which in this case happens to be from limestone thin micrograph sections. Interestingly, even after including a human class as part of the training, the algorithm predicts this photo as human with 86% confidence (Figure 1b). This is most likely due to the grayscale color and some of the background texture, which are very similar to examples in the thin micrograph limestone section training class (Figure 1c).

Another curious case of misclassification is shown in Figure 1d. The figure is a burial history curve from Nunn (2012) and is commonly used to understand source rock genesis and timing, but the model classifies the image as a velocity model with a 48% probability or, secondarily, as a burial history curve (46% probability). Figure 1e shows a burial history curve of very similar nature to the failed prediction in terms of text content, color, and layout and due to the abundance bias of velocity model color schemes. It is still unable to get it correct. This is due to

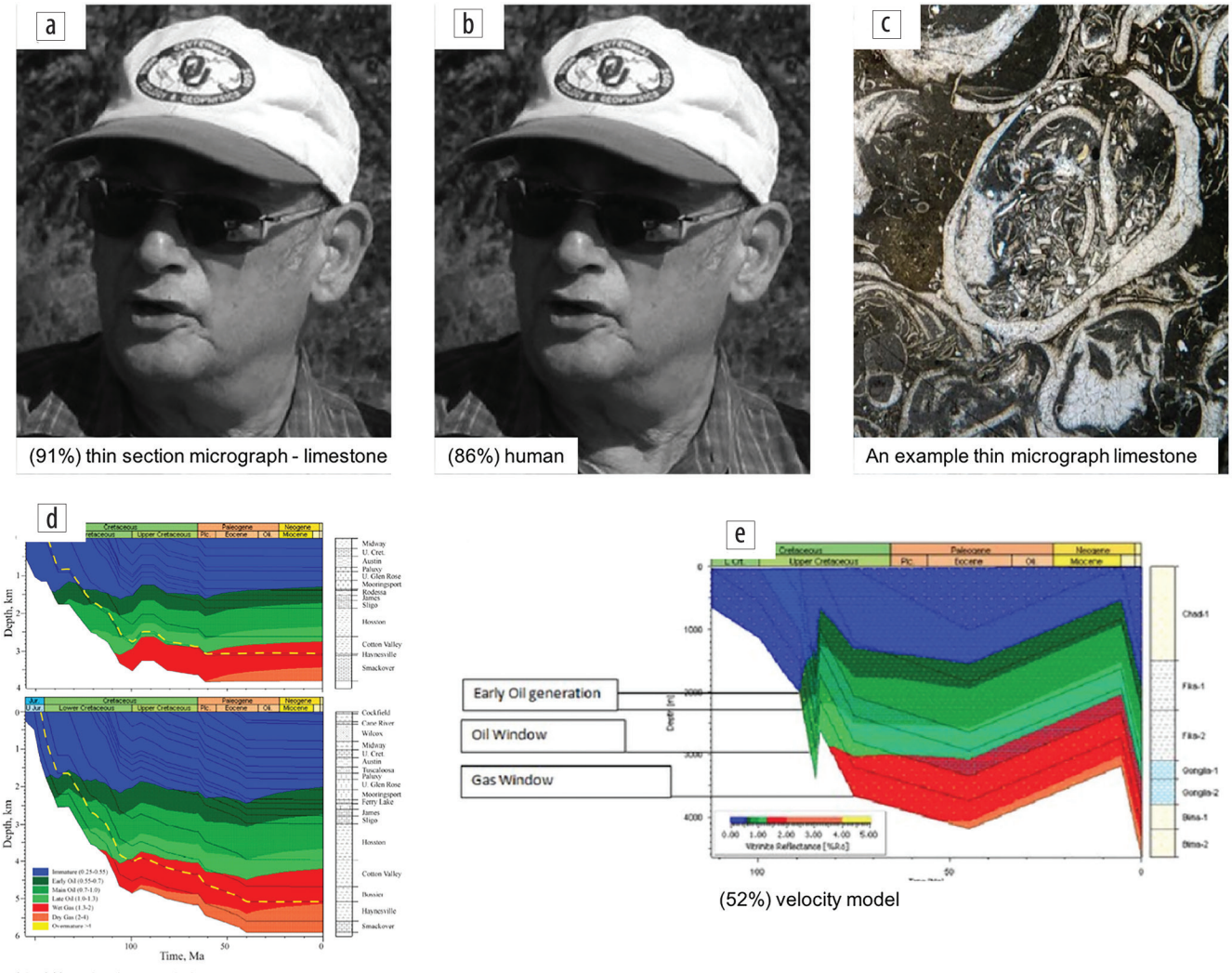

Figure 1. Image segmentation examples. A misclassification of a human photo from a paper by Infante-Paez et al. (2017) with limestone thin section (a, b, and c). A similar example for burial history misclassified as a velocity model from a paper by Nunn (2012).

various factors that the machine learning algorithm fails to characterize correctly. The first issue is that the color schemes are very similar to generic velocity models, which by itself is an individual class in the algorithm. The second issue is that without other context in the image, such as text and placement of text, the classifier cannot differentiate between these two classes. Lastly, the abundance of velocity models with a similar layout and color scheme (approximately 30) with respect to the abundance of similar color and layout of burial history curves (approximately 5) lends a strong bias to the classifier. This can be remedied by taking care when building the training set. For similar feature layouts in image schemes, one should seek to ensure that appropriate populations of similar images across classes are represented to achieve good separation in the prediction space.

These examples clearly demonstrate that there is no free lunch in machine learning applications. The training data set needs to be robust for the predictions to perform accurately. It also shows the possibility of learning from the outcomes of machine learning algorithms. In the following sections, we show how similar scenarios can happen in more familiar geoscience applications.

Well-log-based facies classification

In this section, wells from the Forties Field are used to examine the performance of various machine learning algorithms for facies classification. The Forties oil field, located 110 mi (177 km) offshore Aberdeen within the UK production block 21/10 at a water depth of 106 m, is considered to be one of the oldest and largest oil fields in the UK North Sea with estimated oil reserves in place of 5 billion barrels. Discovered in 1970 by BP, the field was brought online in 1975. Forties is a classic example of enhanced recovery drilling assisted by time-lapse seismic studies. This has extended the life expectancy of the Forties by 20 years, and after around 40 years of production, the field continues to produce at production rates of approximately 45,000 BOPD. The Forties Formation consists of a lower shale member and an upper sandstone member, which were deposited in a middle and lower submarine fan environment.

Twenty-nine key exploration and appraisal wells were selected for this study. The input logs for facies classification are compressional and shear sonic logs, density, volume of shale, and porosity logs. The wells were drilled over the early production period. The experiment conducted here is such that the performance of different algorithms and the impact of data abundance can be analyzed thoroughly. To achieve this, two different scenarios are performed in which scenario 1 employs six wells for training the classifier, and scenario 2 employs 27 wells. In each scenario, 80% of the data is used for training, and 20% is left to measure and track accuracy during training. Furthermore, two wells are kept as blind wells to perform cross validation.

Figure 2 shows one of the blind wells, which allows cross validating the different classifiers for accuracy. It is also possible to compute the accuracy when the two data size scenarios are used for training. Once the algorithm is trained, the application to the new wells takes seconds. It can be observed that all classifiers have a reasonable prediction except at the zone marked with a black

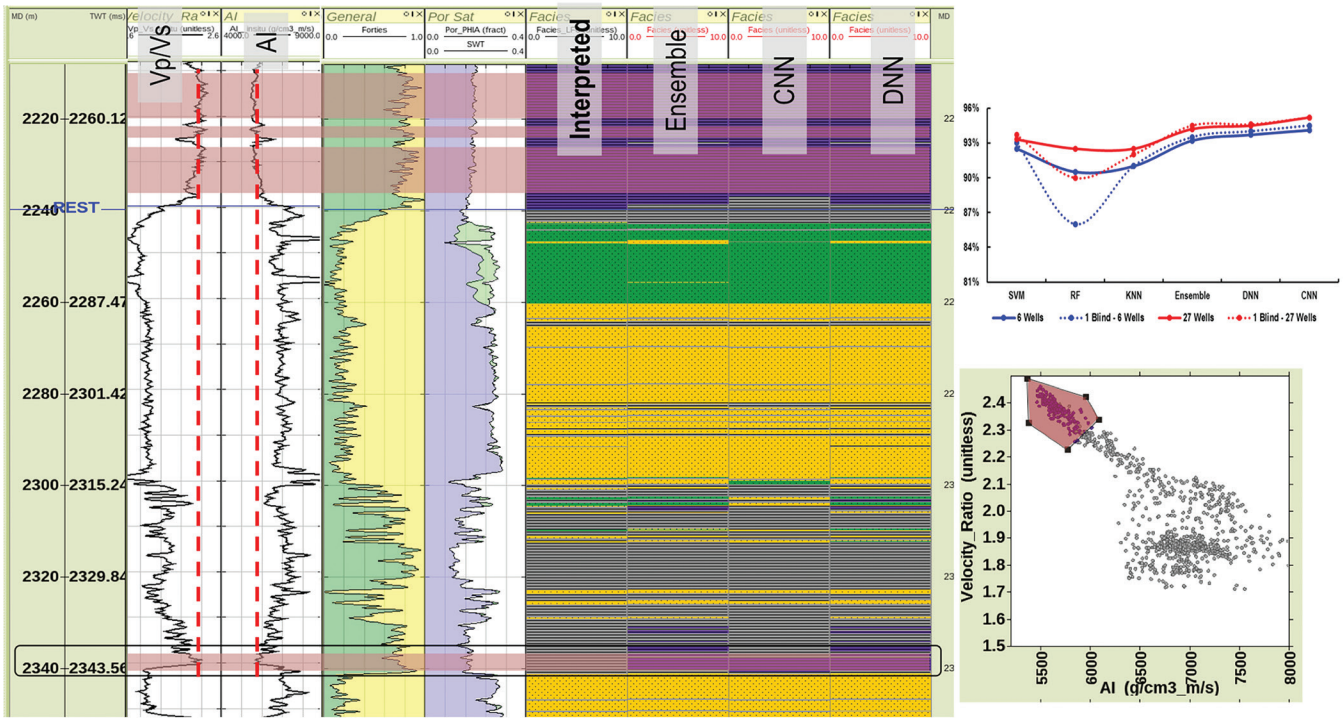

Figure 2. Well-log panel summarizing the result of various classifiers. As expected, the accuracy is higher when more data are used for training (top right) — the solid and dotted lines show the training and validation accuracy. SVM: support vector machines. RF: random forests. KNN: k-nearest neighbor. Ensemble: an ensemble of SVM, RF, and KNN. CNN: convolutional neural network. DNN: deep (multilayer) neural network. The polygon data selection on the crossplot (bottom right) highlights the corresponding samples on the well-log panel where the similarity of the elastic response is evident between Sele soft shale and what is picked within the black box. The learning from the machine here is a possibility of interpreting a new type of intrareservoir shale. This could have a big impact on the subsequent reservoir-characterization workflows such as facies-based seismic inversion.

rectangle. It appears that all classifiers predict soft shale facies for this zone. It is geologically impossible to have soft shale from the Sele Formation at the deeper formations; however, this may indicate that the interpretation can be modified to include a new type of intrareservoir shale facies. This can be explored by analyzing the acoustic impedance (AI) and V_P/V_S logs as shown in Figure 2. The red-dashed lines show that AI and V_P/V_S have similar values within the black rectangle and the shallower Sele soft shale, which could be envisaged as a possibility of interpreting a new type of shale within the reservoir zone. Note that if one is confident that such facies is unnecessary to have at that depth then depth itself can be used as a feature in the training, which would guide the classifiers to produce something more similar to the interpreted facies. However, the point is that QC for the machine learning algorithm has to be done with care, and the outcomes can be seen as a way to learn from machines. (This is somewhat equivalent to human misclassification in the previous section.)

Petrophysical log estimation

Broadly speaking, the inputs to petrophysics workflows are gamma ray, deep resistivity, density, neutron, and sonic logs, and the main outputs are volume of shale, effective porosity, and water saturation. Hopefully it is clear to readers that the outputs are continuous properties, which means the problem is suited for regression (as opposed to classification discussed in the previous section). In machine learning, classification and regression algorithms are similar except for the output (target) variables. In classification, the target is a discrete property (e.g., a set of predefined facies identified with integer numbers as in the previous section). In petrophysics prediction, one aims to model a continuous log using regression (i.e., volume of shale and porosity). This ultimately means one can use most classification algorithms for regression.

It is important that the input data are conditioned to remove any instances of erroneous log response, e.g., logs responding to poor hole conditions or logs acquired through casing. For this experiment, data from the Central North Sea area were used. In each well, the petrophysics logs (required for training and testing) have been derived using standard deterministic petrophysics methodology. Four wells were used as a training data set, and one was used as a blind well test in a similar QC strategy as in the previous section with the objective to compute the volume of shale and porosity. Hence, gamma ray, density, resistivity, and neutron logs from the four wells were used (again 80%–20% data split for training and testing) to predict volume of shale. Then given the volume of shale, the same process was repeated to predict porosity. The outcome is shown on the blind well in Figure 3. Overall, it can be observed that the prediction is good for both volume of shale and porosity logs (total root-mean-square error is 0.02). However, a closer inspection (Figure 3b) of the sand-dominated zone (yellow area in density log) indicates some error between the interpreted (black) and predicted porosity (red) logs. After analyzing the training data in more detail, the learning from the machine here is that the training data set did not have sandstone with such high porosity values — yet another example of the impact of data completeness on the accuracy of machine learning algorithms. This example shows how one can trace the inconsistency observed in the outcome of machine learning. Having more data helps remedy this problem, and like other examples, once a satisfactory training quality is obtained, the application to new wells is fast. Having more data and expecting more wells is the ideal scenario (e.g., a regional study) that such automatic petrophysical interpretation pays off. (A companion paper by Zabihi Naeini et al., scheduled for the January 2019 issue of *TLE*, will show an example of using 30 wells from the central North Sea for automatic petrophysical interpretation).

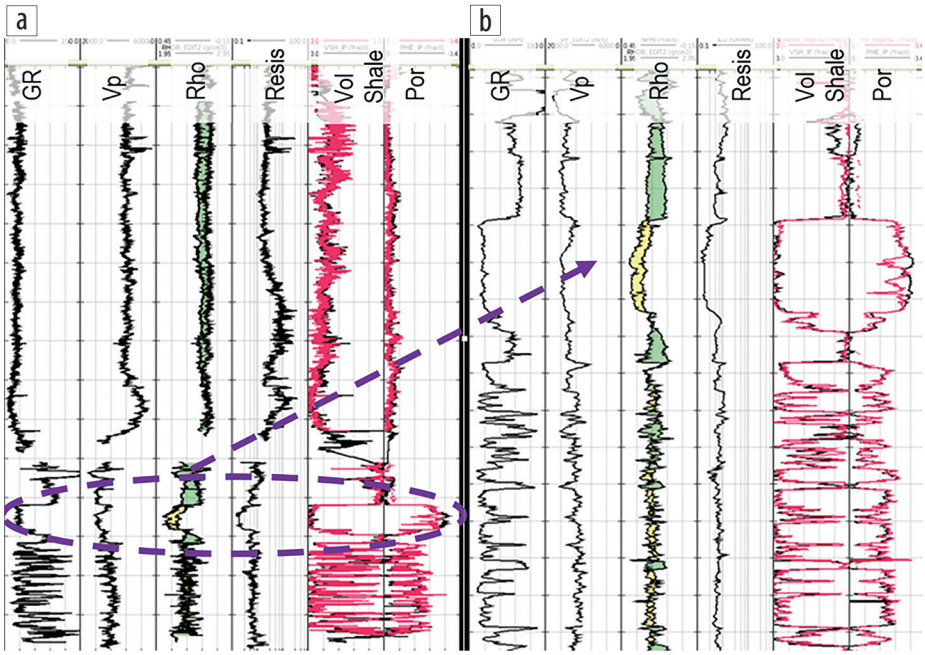

Figure 3. (a) Volume of shale and porosity prediction for a blind well. (b) Zoom over the sandy section in ellipse. The interpreted and predicted logs are shown in black and red.

Automatic fault prediction

Fault interpretation has long been an active topic in the domain of 3D seismic interpretation and reservoir characterization. Since the introduction of the coherence attribute by Bahorich and Farmer (1995), geoscientists have devoted much time to developing new seismic attributes, methods, and algorithms to help detect, depict, and extract faults of interpretational interest from surrounding nonfaulting features (Di and Gao, 2016). For example, the most popular attributes are coherence (Bahorich and Farmer, 1995), semblance (Marfurt et al., 1998; Tingdahl and De Rooij, 2005), curvature (Roberts, 2001), flexure (Di and Gao, 2017; Qi and Marfurt, 2017), homogeneity, and more of such derivatives (e.g., Luo et al., 1996; Gersztenkorn and Marfurt, 1999; Cohen and

Coifman, 2002; Di and Gao, 2014; Wang et al., 2016; Qi and Marfurt, 2017). With the aid of one or more seismic attribute maps, the faults can be extracted assuming two geologic constraints: (1) the fault geometry is continuous and (2) the fault plane is planar or slightly bent (e.g., listric faults). Such fault interpretation tools include ant tracking (Pedersen et al., 2002), Hough transform (AlBinHassan and Marfurt, 2003), principal component and eigenvector analysis (Barnes, 2006), dynamic time warping (Hale, 2013), motion vector (Wang et al., 2014), and more. However, all these methods utilize and parse only a limited number of attributes at a time and consequently are highly dependent on the quality of the used seismic attributes. This limitation can be resolved with the support of emerging machine learning techniques, which are capable of incorporating and identifying the fault in a high-dimensional feature domain. Among these machine learning techniques, artificial neural networks are the most popular, owing to their variable forms of network architecture. For example, deep neural networks consist of multilayers of fully connected neurons and fundamentally rely on a set of attributes (e.g., edge, sharpness, amplitude, etc.) provided as inputs. Convolutional neural networks have given rise to exceptional accuracy in image classification and segmentation in which the network itself generates the required attributes via multiconvolutional layers acting as filters on the image. Convolutional neural networks also have multilayer neurons similar to deep networks. Our experimentations show that when the quality of the seismic data is good, and where the discontinuities are preserved, both the convolutional and deep neural networks give similar accuracy, but convolutional neural networks can perform slightly better and save the user time when providing various attributes. However, there has not been much research on how well the convolutional and deep neural networks compare when the data are challenging with complex structure and amplitude signatures.

The data set utilized for this study is a small part of a large (approximately 3000 km^2) 3D prestack depth migration seismic survey situated in the Browse Basin offshore North West Australia. The Browse Basin is a northeast-trending Paleozoic to Cenozoic depocenter covering an area of approximately 140,000 km^2 with sedimentary successions in excess of 15 km thick. There have been numerous discoveries (Poseidon, Crown, Zephyros, etc.) in the region and multiple resultant field developments including Ichthys, Torosa, and Prelude. The primary exploration target interval is the Plover Formation, which is comprised of synrift fluvio-deltaic sediments of early to middle Jurassic age. Depositional processes exert a primary control on reservoir quality with the best reservoir quality exhibited by tidal and/or tidally influenced channel sequences (Tovaglieri and George, 2012). A major complicating factor in this basin is the presence of extensive volcaniclastic and extrusive igneous deposits relating to active volcanism during the time of deposition. This gives rise to complex seismic amplitude geometries

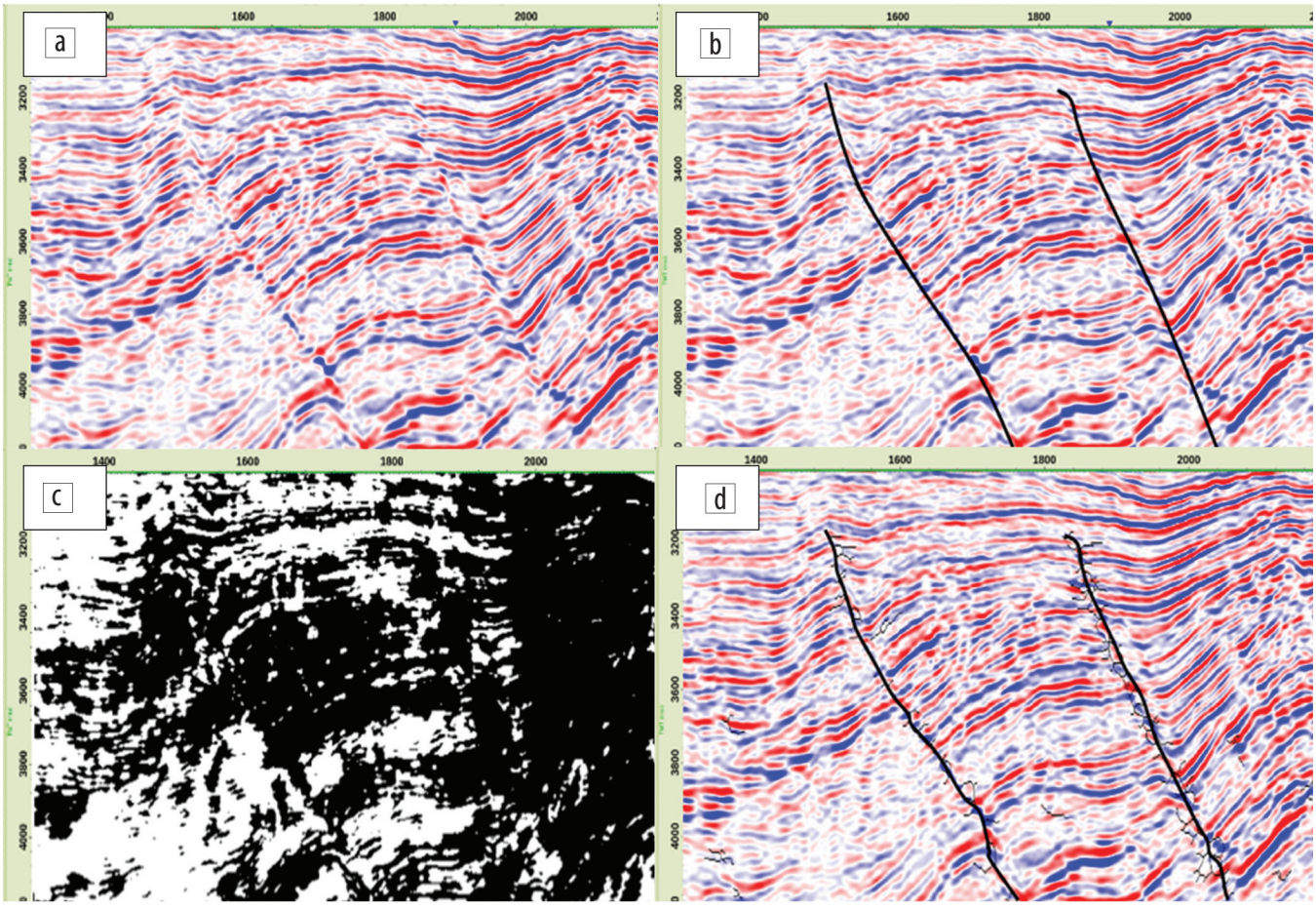

Figure 4. (a) Seismic image, (b) interpreted faults for training, (c) faults from deep neural network, and (d) from convolutional neural networks.

that span much of the Permian/Jurassic interval. A further complication is the complex burial history, which saw multiple phases of extension, thermal subsidence, and inversion. This gives rise to significant faulting and numerous unconformities through the Permian-Jurassic sequences, making identification of major boundaries problematic. Seismic mapping of the Plover Formation is therefore extremely challenging.

A seismic section from this data set is shown in Figure 4a. The complex structure and lack of a sharp discontinuity across the fault planes is evident. The two major faults are shown in Figure 4b. As mentioned earlier, dip, discontinuity, and homogeneity were selected as the input seismic attributes for the deep neural network and the training was performed on a handful of sections (not shown here). During this study, it became clear that none of these attributes are capable of clearly highlighting the major faults. This is due to the presence of seismic noise, a lack of resolution, and the complexity (potentially related to the imaging) of the faults. Hence, it is of no great surprise that the deep neural network is unable to detect the faults as shown in Figure 4c. On the other hand, the convolutional neural network followed by a thinning mechanism provides a satisfactory image of the faults as shown in Figure 4d. It is also interesting that the algorithm is able to detect some of the smaller faults (but also one could argue that it misses some of the smaller ones too) shown with thinner lines, some of which could be interpreted as artifacts. The learning from machines is that in cases where one is unable to generate sufficient attributes for image feature detection (such as faults) either due to the lack of time, expertise, or appropriate software, it is probably safer to let convolutional neural networks figure out what is needed to optimally achieve such tasks. There is no doubt that Figure 4d is more representative of the faults in Figure 4b than what we observe in Figure 4c.

Computational complexity

Most machine learning algorithms in two dimensions and/or single trace algorithms do not require special handling for computation. Most are capable of running on a laptop or a local workstation. On the other hand, understanding the computational complexity of 3D multiensemble deep neural networks is still a developing research theme and is likely to change in the coming years as algorithms, model, and data parallelization techniques are further developed to solve larger problem types. Thus far, the problems solved have been predominantly, but not limited to, computer vision and natural language processing. Recent deep neural networks have been applied pervasively to medical imaging (Greenspan et al., 2016) and most recently to seismic images. This paper is not meant to focus on computational challenges, but we will attempt to review a few metrics and complications that arise from distributed and deep machine learning algorithms. As an analog, a reference scanning electron microscope (SEM) scan library contains approximately 1000^3 pixels for the whole SEM volume, which is similar to an average postmigration seismic cube (although modern seismic acquisition and processing techniques produce denser cubes). A typical 3D 100-epoch run contains hundreds of thousands (if not millions depending on the objective function descent rate and target) of 3D patch extractions and convolutions at multiple patch sizes ($256 \times 256 \times 256$, $128 \times 128 \times 128$, e.g., Figure 5, $64 \times 64 \times 64$ in this example). Although a single convolution is much faster than a single trace migration, the shear multitude makes them more computationally intensive over the whole deep neural network. Figure 5 shows an image of such a 3D cube resulting from a super resolution generative adversarial network (GAN). GANs are a special type of algorithm consisting of generator and discriminator networks where the generator attempts to fool the discriminator into thinking it is seeing a valid model (Goodfellow et al., 2014). A test run using model parallelization took roughly 1000 GPUs approximately 24 hours to segment and infer a 1000^3 sample seismic volume. Other algorithms, such as classic convolutional networks with shallower structure, can take less time to converge, reducing run time to just an hour or two on 30 GPUs. The fault extraction example in the previous section, with three convolutional layers and two fully connected layers training on four inlines, took approximately 3 hours on a laptop with eight cores, which reduces to about half an hour on GPUs.

Fortunately, there is more good news. Deep neural networks, once trained, can perpetually make predictions in a fraction of time. Furthermore, with good-quality data that are usable for training and by means of transferred learning, the model can only become better in accuracy.

Conclusions

While implementing various machine learning applications, it became apparent that quality controlling the outcomes plays an important role not only in building confidence in the algorithm but also in addressing two skepticisms: (1) the concern that machines will replace humans and (2) concern over black-box-type algorithms. While both concerns are valid in the abstract, the concept of machine learning and learning from machines helps address them in real-life problems. This makes machine learning algorithms enablers (rather than replacers) for practitioners to help do various tasks faster, and they could also help improve human interpretations and/or highlight human errors. Various examples were shown here. In practice, these things are virtually always observed. Ultimately, the interaction between machine and human can be implemented such that the hybrid outcome is the most optimized. TLE

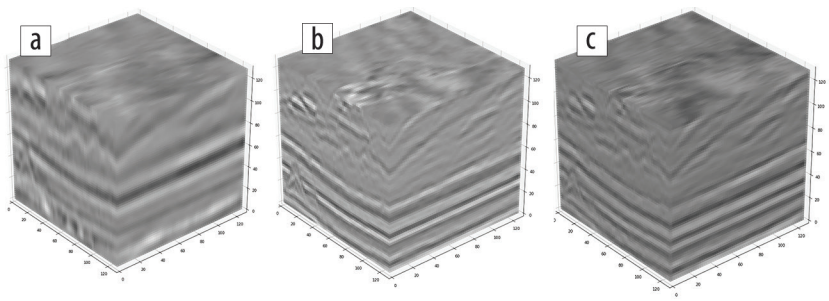

Figure 5. (a) Low-frequency seismic image, (b) ground truth seismic, and (c) predicted result from a GAN.

Acknowledgments

We would like to thank Haibin Di for useful discussions on fault prediction.

Corresponding author: enaeini@ikonscience.com

References

AlBinHassan, N. M., and K. Marfurt, 2003, Fault detection using Hough transforms: 73rd Annual International Meeting, SEG, Expanded Abstracts, 1719–1721, https://doi.org/ 10.1190/1.1817639.

Bahorich, M., and S. Farmer, 1995, 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube: The Leading Edge, **14**, no. 10, 1053–1058, https://doi.org/10.1190/1.1437077.

Barnes, A. E., 2006, A filter to improve seismic discontinuity data for fault interpretation: Geophysics, **71**, no. 3, P1–P4, https://doi.org/10.1190/1.2195988.

Bayes, T., and M. Price, 1763, An essay towards solving a problem in the doctrine of chances: Philosophical Transactions, **53**, 370–418, https://doi.org/10.1098/rstl.1763.0053.

Cohen, I., and R. R. Coifman, 2002, Local discontinuity measures for 3-d seismic data: Geophysics, **67**, no. 6, 1933–1945, https://doi.org/10.1190/1.1527094.

Cortes, C., and V. Vapnik, 1995, Support-vector networks: Machine Learning, **20**, no. 3, 273–297, https://doi.org/10.1007/BF00994018.

Crevier, D., 1993, AI: The tumultuous search for artificial intelligence: Basic Books.

Deng, J., W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, 2009, ImageNet: A large-scale hierarchical image database: Conference on Computer Vision and Pattern Recognition, IEEE, https://doi.org/10.1109/CVPR.2009.5206848.

Di, H., and D. Gao, 2014, Gray-level transformation and Canny edge detection for 3D seismic discontinuity enhancement: Computers & Geosciences, **72**, 192–200, https://doi.org/10.1016/j.cageo.2014.07.011.

Di, H., and D. Gao, 2016, Seismic attribute-aided fault detection in petroleum industry: A review, *in* D. Martin, ed., Fault detection: Methods, applications and technology: Nova Science Publishers, 53–80.

Di, H., and D. Gao, 2017, 3D seismic flexure analysis for subsurface fault detection and fracture characterization: Pure and Applied Geophysics, **174**, no. 3, 747–761, https://doi.org/10.1007/s00024-016-1406-9.

Fukushima, K., 1980, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position: Biological Cybernetics, **36**, no. 4, 193–202, https://doi.org/10.1007/BF00344251.

Gersztenkorn, A., and K. J. Marfurt, 1999, Eigenstructure-based coherence computations as an aid to 3-d structural and stratigraphic mapping: Geophysics, **64**, no. 5, 1468–1479, https://doi.org/10.1190/1.1444651.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, 2014, Generative adversarial networks: Proceedings of the 27th International Conference on Neural Information Processing Systems, **2**, 2672–2680.

Greenspan, H., B. van Ginneken, and R. M. Summers, 2016, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique: Transactions on Medical Imaging, IEEE, **35**, no. 5, 1153–1159, https://doi.org/10.1109/TMI.2016.2553401.

Hale, D., 2013, Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images: Geophysics, **78**, no. 2, O33–O43, https://doi.org/10.1190/geo2012-0331.1.

Ho, T. K., 1995, Random decision forests: Proceedings of the 3rd International Conference on Document Analysis and Recognition, IEEE, 278–282 https://doi.org/10.1109/ICDAR.1995.598994.

Hochreiter, S., and J. Schmidhuber, 1997, Long short-term memory: Neural Computation, **9**, no. 8, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Hopfield, J. J., 1982, Neural networks and physical systems with emergent collective computational abilities: Proceedings of the National Academy of Sciences of the United States of America, **79**, no. 8, 2554–2558, https://doi.org/10.1073/pnas.79.8.2554.

Infante-Paez, L., L. F. Cardona, B. McCullough, and R. Slatt, 2017, Seismic analysis of paleotopography and stratigraphic controls on total organic carbon: Rich sweet spot distribution in the Woodford Shale, Oklahoma, USA: Interpretation, **5**, no. 1, T33–T47, https://doi.org/10.1190/INT-2015-0151.1.

Jones, K. S., 1973, Index term weighting: Information Storage and Retrieval, **9**, no. 11, 619–633, https://doi.org/10.1016/0020-0271(73)90043-0.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, ImageNet classification with deep convolutional neural networks: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097–1105.

Laplace, P., 1812, Théorie analytique des probabilités: Courcier.

LeCun, Y., C. Cortes, and C. Burges, 2018, The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/, accessed 8 October 2018.

Legendre, A. M., 1805, Nouvelles méthodes pour la détermination des orbites des comètes: Firmin Didot.

Linnainmaa, S., 1970, The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors: M.S. thesis, University of Helsinki.

Luo, Y., W. Higgs, and W. Kowalik, 1996, Edge detection and stratigraphic analysis using 3D seismic data: 66th Annual International Meeting, SEG, Expanded Abstracts, 324–327, https://doi.org/10.1190/1.1826632.

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**, no. 4, 1150–1165, https://doi.org/10.1190/1.1444415.

Markov, A. A., 1906, Extension of the law of large numbers to dependent quantities [in Russian]: Izvestiya Fiziko-Matematicheskikh Obschestva Kazan University, **15**, 135-156.

McCarthy, J., and E. Feigenbaum, 1990, Arthur Samuel: Pioneer in machine learning: AI Magazine, **11**, no. 3, 10–11.

Nunn, J. A., 2012, Burial and thermal history of the Haynesville Shale: Implications for overpressure, gas generation, and natural hydrofracture: GCAGS, **1**, 81–96.

Pedersen, S. I., T. Randen, L. Sonneland, and Ø. Steen, 2002, Automatic fault extraction using artificial ants: 72nd Annual International Meeting, SEG, Expanded Abstracts, 512–515.

Qi, X., and K. Marfurt, 2017, Volumetric aberrancy to map subtle faults and flexures: 87th Annual International Meeting, SEG, Expanded Abstracts, 3443–3447, https://doi.org/10.1190/segam2017-17632807.1.

Roberts, A., 2001, Curvature attributes and their application to 3D interpreted horizons: First Break, **19**, no. 2, 85–100, https://doi.org/10.1046/j.0263-5046.2001.00142.x.

Rosenblatt, F., 1958, The perceptron: A probabilistic model for information storage and organization in the brain: Psychological Review, **65**, no. 6, 386–408, https://doi.org/10.1037/h0042519.
Rumelhart, D., G. Hinton, and R. Williams, 1986, Learning representations by back-propagating errors: Nature, **323**, 533–536, https://doi.org/10.1038/323533a0.
Tingdahl, K. M., and M. De Rooij, 2005, Semi-automatic detection of faults in 3D seismic data: Geophysical Prospecting, **53**, no. 4, 533–542, https://doi.org/10.1111/j.1365-2478.2005.00489.x.
Tovaglieri, F., and A. George, 2012, Sedimentology and image-log analysis of the Jurassic Deltaic Plover Formation, Browse Basin, Australian North West Shelf: Presented at Annual Convention and Exhibition, AAPG.
Turing, A., 1950, Computing machinery and intelligence: Mind, **59**, 433–460.
Wang, S., S. Yuan, B. Yan, Y. He, and W. Sun, 2016, Directional complex-valued coherence attributes for discontinuous edge detection: Journal of Applied Geophysics, **129**, 1–7, https://doi.org/10.1016/j.jappgeo.2016.03.016.
Wang, Z., Z. Long, G. AlRegib, A. Asjad, and M. A. Deriche, 2014, Automatic fault tracking across seismic volumes via tracking vectors: International Conference on Image Processing, IEEE, 5851–5855.
Watkins, C. J., 1989, Learning from delayed rewards: Ph.D. thesis, University of Cambridge.