Redefining seismic interpretation - Machine learning for fault interpretation, enhancing efficiency, accuracy and auditability through a cloud-based approach *

Surender Manral ***, Francis Grady **, Olesya Zimina *** and Guido van der Hoff **

(Received September 19, 2018 : accepted November 6, 2018)

Abstract: Market dynamics have challenged the oil and gas industry to evolve. Modern digital technology has started to impact the entire oilfield lifecycle from exploration to development and production, fundamentally changing the way geoscientists work by enhancing performance and enabling significant value creation.

In subsurface characterization, computer-assisted seismic interpretation has been around for several decades. Over time, computer and software technology advances have improved the speed and quality of seismic interpretation, but these advances have coincided with an exponential growth in the volume of data to be interpreted. Consequently, the critical task of performing seismic interpretation is both time-consuming and laborious. Moreover, friction in accessing data and relevant technologies, along with a lack of insight due to inefficient collaboration, increases the uncertainty of the results. Due to these factors, it takes several months to mature an interpretation and build a 3D digital representation of the subsurface which is required to support a drilling decision.

This paper is primarily focused on seismic fault identification, which is a key component in subsurface characterization and modelling workflows, using a combination of cloud technology and new machine learning techniques.

Keywords: fault interpretation, machine learning, labels, fault prediction

1. Introduction

Accurate fault interpretation is a key step in subsurface characterization studies. It enables the identification of drilling hazards, minimizing the risk associated with well placement and is a critical input into subsurface reservoir characterization studies such as basin evolution, petroleum systems modelling, and reservoir, trap seal analysis studies. Traditionally, fault interpretation is predominantly a manual and time-consuming task that is highly dependent on the skill of the interpreter. As seismic interpretation is knowledge and experience intensive, obtaining an accurate interpretation requires significant expertise from the geoscientist, which may take several years or more to acquire. One of the examples where specialized skills are necessary is the generation and use of seismic attributes. For many years, seismic attributes have been used as a guide by interpreters for the task of mapping faults. However, with the number of attributes available and the daunting parametrization associated with each attribute, it may take a skilled user hours to days to explore and identify the best fault attribute for a given dataset. In general, to build a robust fault framework, can take an expert weeks to months of work.

Automation of the manual process can help to reduce the time required for interpretation allowing geoscientists to focus on more important and complex activities. It can also provide an intelligent way of getting expected results with minimum user interaction. Automation in fault interpretation can help in changing the role of an interpreter from one in which they manually, laboriously digitize interpretations to one in which they review and accept or reject an interpretation. A first step in this process is to provide parameterless, automated interpretation guidance usable by any geoscientist. Machine learning along with elastic cloud compute has tremendous potential to significantly improve efficiency in fault interpretation workflows whilst providing greater insight into expensive and unwieldy seismic datasets.

2. Method

In the software industry, there has been a trend towards
data-driven machine learning (ML) models. ML is a type of artificial intelligence in which computers trained to recognize patterns without being explicitly programmed. After a model has been suitably trained, it can then be used to predict similar patterns in data it has never seen before. There are a variety of fields in which ML models are used, such as credit card fraud detection, speech recognition and equipment failure prediction.

For fault interpretation, a novel ML approach was designed. The goal of this design is to provide interpreters with guidance on fault locations without requiring user-supplied parameters. Fig. 1 shows the back-end architecture diagram for the training of the ML for fault interpretation and the fault prediction service.

One of the most critical components of the whole ML ecosystem is the selection of a specific ML model. Computer science journals and papers (Johnston et al., 2017; Meskó et al., 2017) suggest many different model architectures,

![Architecture overview for the fault identification through ML](image)

**Fig. 1** Architecture overview for the fault identification through ML

a) Overview of the fault training; b) fault prediction

![Machine learning in biological imaging: Supervised retinal vessel segmentation](image)

**Fig. 2** Machine learning in biological imaging: Supervised retinal vessel segmentation (Memari et al., 2017)
while machine learning competitions, such as the ImageNet Challenge, promote even more options. To identify a suitable architecture for fault identification, the choice of multiple ML solutions has been narrowed to biological imaging models (Fig. 2). These models are often used to process data with similar challenges to those faced by seismic interpreters, therefore a model that performed well in several biological imaging tasks, was identified and adapted to better fit seismic input data.

After the model architecture was selected, the next step was to train an ML model to identify faults. Training the model required a wide variety of seismic data along with expertly labelled fault interpretations. Training labels were provided, both their accuracy and completeness of the interpretations are paramount to ensure that the model is not incorrectly trained. To collect these quality data, several seismic datasets have been used, along with a group of seismic interpreters with 10 to 15 years of fault interpretation expertise. Also, for traceability and future data analytics purposes, a data repository structure and a metadata tagging system were established.

After the fault interpretation data, which in this case are called training data, were collected, the next step was preprocessing. During this step, the data were cleaned, sampled and converted into an easily readable format and applied as input into the training together with the predefined ML model (Fig. 1a).

A challenge with training any ML model is the iterative nature of the training process itself. The model must be trained on the data hundreds or thousands of times. Together with the large size of seismic datasets to complete training would take weeks or months on a standard workstation CPU. To address this issue the training process was systematically reviewed and all components optimized. Graphical processing units (GPUs) were used to further speed up the training. Finally, to make use of more compute resources than were locally available, training was run on commercially provided Google Cloud GPU’s. This provides a key advantage in overall performance comparative to the traditional approach.

The output of the training pipeline is the ML model/brain, which then can be applied on a new seismic dataset that the ML model has not seen during training. The ML model can help automatically to recognize and highlight faults and faulting patterns, which is a process known as “prediction” (Fig. 1b).

The whole ML ecosystem is designed such that the prediction result can be enhanced by retraining the model if required. To retrain, a geoscientist provides new training data that represents several fault interpretations delineated on the seismic cube used for the prediction exercise. These additional training data are used as input to retrain and tune the existing ML model, further refining the prediction quality.

One of the key components of our approach is accessibility and the frictionless experience from an end-user perspective. As the data is on the cloud, it can be accessed from anywhere and at any time. A geoscientist can run fault prediction almost instantaneously by leveraging the elastic compute power of the cloud without launching many dialogs and fine tuning the parameters.

3. Results

The ML fault model used for the prediction of the faults in this paper was trained on 10 to 12 datasets from a variety of geological basins around the world. This fault model was primarily trained for identifying major faults, because most of the minor faults were not labeled in the training data. The dataset used in this paper for prediction was a data set from the northwest shelf of Australia that had never been seen by the ML fault model.

The ML fault prediction model accurately predicts most of the major fault locations. It not only is able to identify obvious faults, in some other cases our network also identified non-obvious faults that would be challenging to characterize using seismic attributes. Comparing our ML results to a classical structural attribute generated with default parameters (Fig. 4a), the structural attribute shows the discontinuities, but results are comparatively noisy and noncontinuous. We can also observe clinoforms in the structural attribute running perpendicular to the major fault trend. In contrast, the ML fault prediction results are continuous and relatively free of

Fig. 3 Schematic for the proposed end-user fault prediction workflow
noise (Fig. 4b).

We can clearly observe most of the major faults are picked with precision by the ML fault model (Fig. 5). In Fig. 5b, the green arrow shows faults predicted in an area of low S/N. Red arrows show some of the faults missed by the fault model.

To enhance the identification of minor faults, the ML model can be further trained on new data containing subtle or minor faults, to provide an updated prediction.

Automatic extraction and segmentation of the identified faults is an obvious next step in this process. The continuity and low noise of the fault prediction results make them ideal for an automated extraction and segmentation process that generates fault planes with minimal user effort.

4. Conclusion

Through the novel use of ML, we have successfully created a system for automated, parameterless seismic fault interpretation guidance. Using correctly labeled training data,

![Fig. 4 Time slice](image)

a) Structural attribute with default parameter. b) Overlay of structural attribute with the ML fault prediction (with opacity). Yellow color shows the high probability of the faults.

![Fig. 5 The inline intersection](image)

a) Seismic amplitude. b) Overlay of seismic amplitude with the ML fault prediction (showing high probability of faults using opacity)
the ML model was trained using a framework running on cloud compute resources. The resulting ML brain is used to create fault interpretation guidance, by providing predictions of fault likelihood anywhere within a seismic volume. The solution presented in this paper provides a strong alternative to conventional fault detection methods using seismic attributes, which relies mainly on the experience of the interpreter and can be a human-intensive task. This approach builds upon the expertise of many geoscientists, stored in the ML model, to automatically detect the fault locations.

The ML model prediction is the first step towards a paradigm shift in the way geoscientists conduct seismic interpretation. The approach provides greater efficiency and geological insight as part of routine fault interpretation activity, whilst reducing the subjectivity bias present in manual seismic image analysis.

Acknowledgements

We would like to thank Schlumberger for allowing us to publish this work. Also, we would like to thank Geoscience Australia for the dataset.

References

Johnston D.H., Dorn G., Fomel S., Lomask J., Roth M., and Star T., 2017: Introduction to Special Section: Computer-Assisted Seismic Interpretation Methods. Interpretation, 5 (3), SJi–Sji. https://doi.org/10.1190/int-2017-0615-spscintro.1.

Meskó B., Drobni Z., Bényei É., Gergely B., and Győrfy Z., 2017: Digital Health is a Cultural Transformation of Traditional Healthcare. mHealth 3:38.

Memari N., Ramli A. R., Saripan M. I. B., Mashohor S., and Moghbel M., 2017: Supervised retinal vessel segmentation from color fundus images based on matched filtering and AdaBoost classifier. PLoS One., 12 (12), e0188393. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5724901.