AEGC2019 Data to Discovery
Australasian Exploration Geoscience Conference
2-5 September 2019 • Perth, Western Australia
Incorporating the AIG, ASEG, PESA and WABS

Viability of long-short term memory neural networks for seismic refraction first break detection – a preliminary study

Tasman Gillfeather-Clark*
Centre for Exploration Targeting
University of Western Australia
35 Stirling Hwy, Crawley, WA 6009
tasman.gillfeather-clark@research.uwa.edu.au

Eun-Jung Holden
Centre for Exploration Targeting
University of Western Australia
35 Stirling Hwy, Crawley, WA 6009
eun-jung.holden@uwa.edu.au

Daniel Wedge
Centre for Exploration Targeting
University of Western Australia
35 Stirling Hwy, Crawley, WA 6009
daniel.wedge@uwa.edu.au

Tom Horrocks
Centre for Exploration Targeting, UWA
35 Stirling Hwy, Crawley, WA 6009
tom.horrocks@uwa.edu.au

Carlie Byrne
Rio Tinto Resource Evaluation
152-158 St Georges Terrace
Perth, WA 6000
carlie.byrne@riotinto.com

Matthew Lawrence
Rio Tinto Resource Evaluation
152-158 St Georges Terrace
Perth, WA 6000
Matthew.Lawrence@riotinto.com

SUMMARY

Seismic data processing and analysis focuses on identifying the arrival of seismic waves or 'first-breaks'. The identification of the arrival of first breaks is complicated by the variance of recording quality typically found across the dataset. In an exploration setting, models need to be developed and refined multiple times. Picking these first breaks then becomes time consuming, limiting the interpreter to processing their dataset rather than considering the implications of their model. Machine Learning as a field continues to respond to many data centric issues within geoscience. However, the field as a whole continues to grapple with balancing the power of these new techniques against operator expertise and skill.

This paper presents a methodology to identify the first break in seismic refraction data using a Long-Short Term Memory (LSTM) network, which is a recurrent network architecture. I propose one way to delineate between different groups of traces that the operator would intuitively pick differently, by using dynamic time warping to generate a distance matrix of the seismic traces for clustering. This clustering of trace types allows for a more targeted selection of training samples. I conclude with a proposed framework for the integration of operator skill with machine learning speed and repeatability.

Key words: LSTM, neural networks, dynamic time warping, seismic refraction

INTRODUCTION

Seismic refraction is a well-established geophysical survey technique. The method works by detecting the arrivals of an elastic wave at a ground-based sensor at a known distance from a seismic source. For details of its method and processing the reader is directed to Hawkins (1961), and Palmer (1981), respectively. First break detection or 'picking' is a critical stage of data processing for seismic refraction datasets. By defining the first break you allow for the calculation of subsurface velocities and subsequently define the depth of the basement. First break picking automation is by no means a new problem in geoscience, nor is this the first application of machine learning to this problem. Neural networks have been applied to first break picking by McCormack et al. (1993), as well as more algorithmic approaches (Sabbione and Velis, 2010). However, these methods limit the integration of operator interpretation. This is particularly important in cases where trace first arrivals are soft or in cases where signal to noise is high.

Broadly the approaches of seismic first break detection can be partitioned into algorithmic and machine oriented. I use the term machine oriented in this context to stipulate those that implement the generalised machine learning process surmised as; model training and result prediction. To briefly surmise; early methods had issues with wavelet shape changes (Pealdi and Clement, 1972), a series of statistical tests were proposed for detecting first 'kicks' (Hatherly, 1982), though eventually a popular methodology found its way into commercial use which focused on the sudden increase in signal energy (Coppens, 1985). Following these early attempts more targeted work was done to address issues in these methods; some focus on issues arising from picking traces with a low signal to noise ratio (Boschetti et al., 1996), where others refine earlier methods to improve temporal accuracy (Sabbione and Velis, 2010).

Machine oriented methods for refraction data did not see much initial interest in the literature. Early methods utilizing Neural Networks and Fuzzy Logic systems proposed fairly similar methodologies (McCormack et.al., 1993; Chu and Mendel, 1994). More recently interest has continued to grow in the application of recurrent neural networks to the detection of first arrivals in the field of earthquake prediction (Wang et al., 2017; Asim et al., 2018). The detection of first arrivals of earthquake seismogram data is closely related to the study discussed in this paper, though with notable differences in the datasets used. Seismic refraction data by comparison has exceptionally high temporal resolution. This is necessary due to the speeds at which the elastic waves propagate and importantly the close proximity of the source and receiver. It is also important to highlight that in most of these works they display the accuracy of their picking methods on a collection of seismic traces known as a gather.

The dataset used in this preliminary study, is a subset of the data acquired in Haederle et al. (2016). The refraction methodology in their work (termed sparse refraction), acquires three traces for each seismic source. Some example traces from the dataset are displayed in Figure 1. These exemplify the broad types of traces (noisy and clean) within our dataset.

Note that the noisy example in Figure 1 has high frequency noise, and this noise comes in temporally discrete packages.

One of the initial challenges when presenting this data to a neural network; how we convey not only changes within the signal itself but between signals? In this work I explore simple representation learning in shallow neural networks, and I experiment with dynamic time warping to characterise variance across the dataset.

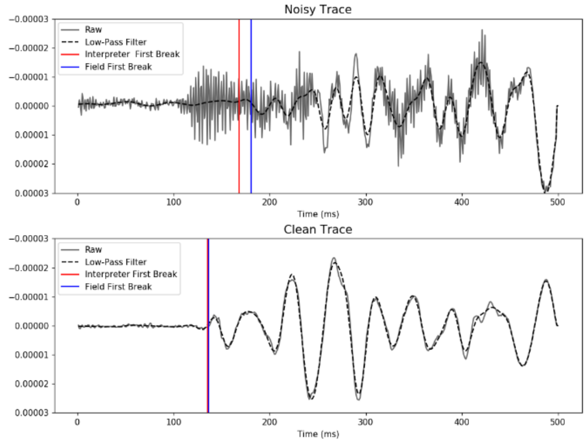

Figure 1. Example of seismic refraction traces from the dataset in use. Note the high frequency noise appears in multiple temporally discrete segments. A low-pass filter has been applied to each trace. There are also the first break picks of each trace. One noted mid acquisition (field) the other chosen in post processing (interpreter).

One of the strengths of the dataset used in this work is the two unique sets of first break picks. The field first break is picked during acquisition. It represents the strongest peak as detected in field conditions, more importantly it is picked on a trace by trace basis. The interpreter pick however reflects a post-processing setting where an interpreter considers the dataset as a whole, typically with the entire survey line of data at their disposal and access to low pass filters. Figure 2 visualises the distribution of the difference between these two picks across the datasets. The majority of operator variance is <40 ms though some traces values can reach approximately 100 ms.

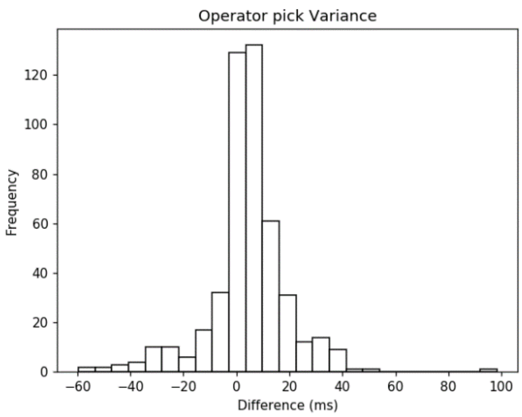

Figure 2. The distribution of the differences between these two operator picks is visualised here. The difference is calculated by subtracting the interpreter pick from the field pick.

METHOD

This study investigates the first break picking on a trace by trace basis. Specifically, we note the stages of pre-processing required train a LSTM network to maximise skill and consistency for first break event detection. It allows the model to learn a representation of the raw and filtered trace and use dynamic time warping (DTW) to characterise changes between traces.

We approach the problem of first break detection as a time-series segmentation problem (Samé et al., 2011). Our seismic signal can quite easily be divided into two categories for classification: no signal (0) and signal (1). Thus, I started by generating a classification label of zero changing to 1 after the interpreter first break occurred.

For the training of neural networks characterising and capturing the variance in trace type is done by regularisation of the training data. Effectively a given sample may statistically represent a small percentage of the overall dataset. However, its features must be learned and retained for later use without dominating all future predictions. In this work I used DTW to generate a distance matrix, which I then used to cluster the traces using hierarchical clustering and Ward's linkage (Johnson and Stephen, 1967) to aid in a representative sampling of the dataset. For more information on dynamic time warping consider Salvador and Chan (2018), and for the package used in this work look to Wannes-Meert, Craenendonck (2018).

Some types of neural networks (Robinson and Fallside, 1987; Williams and Zipser, 1992) also seek to address the problem of overfitting by passing the current hidden state to the next instance, thus integrating past information into the present prediction. However, the values used in these calculations can grow and shrink exponentially. LSTM is an architectural variant of traditional recurrent neural networks (RNN) that overcomes the exponential growth/decay of back propagated errors via the introduction of internal gate units which regulate the passing of information to future model states (Hochreiter and Schmidhuber, 1997).

The gates define what information is retained by a given cell. To understand how they function we need to understand what information is being input into a given LSTM cell including the; cell state, previous hidden state, and input data. There are also two key functions which regulate the cell; these are the sigmoid and the tanh. The sigmoid function output is between 0 and 1, where the tanh function range is between -1 and 1. These regulatory gates take three forms in a *vanilla* LSTM (Greff et al., 2017): forget gates, update gates, and output gates. The forget gate uses the previous hidden state, current input and the sigmoid function to define what information is forgotten. The input gate controls how the current cell state is updated, it takes the current input and previous hidden states, outputs the product of the sigmoid and the tanh functions on these values and adds this to the current cell state. The output gate creates the new hidden state using the output of the tanh function on the current cell state multiplied by the sigmoid function of the old hidden state and the current input. This creates the new hidden state which is passed to the next cell.

Figure 3 visualises the structure of the LSTM network used. Where the LSTM layer is as described above and dense is a fully connected neural layer, where each neuron is connected to another. Its input is the 500 mS long trace each with two features: raw signal and low pass filtered signal. I sampled the dataset using a representative sampling derived from the DTW clustering of the dataset. This ensures that the model retains information on how different traces were picked.

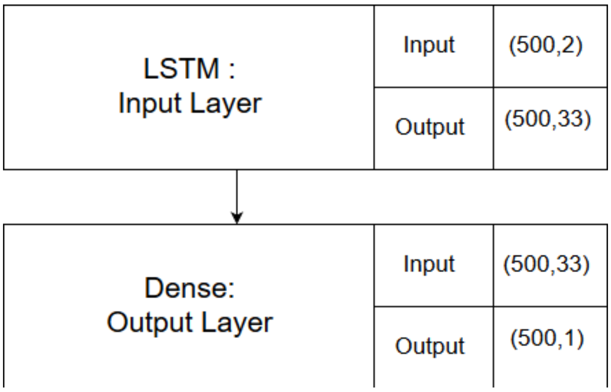

Figure 3. Structure of shallow LSTM neural network used in this work.

The sub selection of the dataset was 1192 traces. The model was trained on 477 traces from this dataset, which took approximately 8 hr. Once trained the model made predictions on 477 traces not in the training data, which took approximately 5 min in total. The remaining 238 traces were set aside for later experimentation.

RESULTS

Before introducing my results, I will outline some simple uncertainty metrics used in this work to evaluate the results and outline DTW and LSTM. I use three methods of uncertainty evaluation, each method is calculated and then normalised to allow for comparison between methods:

- Uncertainty Metric 1 – Operator Variance
- Uncertainty Metric 2 – Noise Components
- Uncertainty Metric 3 – First Break Clarity

Operator variance is the difference between the interpreter and field first break picks; higher values translating to greater uncertainty. Noise components are a comparison of the exact value of the amplitudes of the raw signal compared to the filtered signal. Commonly this would be the signal to noise ratio. In our case, we consider the inverse so that again higher values correspond to greater uncertainty. First break clarity compares the amplitude of the signal 20 mS before and after the first break, where higher values correspond to lower contrast and hence greater uncertainty.

The results of this preliminary study indicate that once trained the model is able to quite accurately detect a first break arrival. This is visualised in Figures 4, 5 and 6, where the distribution of the difference between the model and operator first picks is plotted. Where the difference is negative, the model has picked early, where the difference is positive the model has picked late, and where difference approaches zero it has picked close to the actual value. Each bin is then coloured using the mean value of the uncertainty methods, where higher values correspond to greater uncertainty. This allows for intuitive observations.

Primarily the vast majority of picks show minimal difference between operator and model trained picks. Although, a non-negligible amount, in this instance of the model, does tend to pick later, than operator picks.

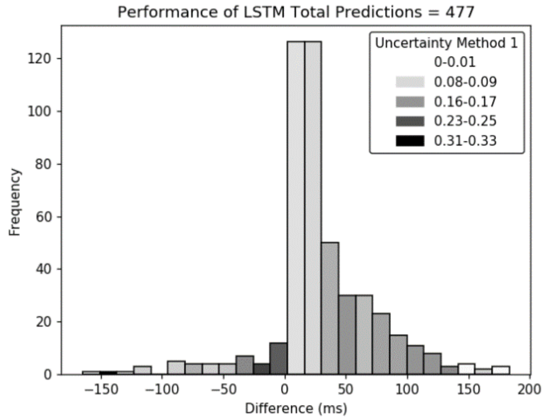

Figure 4. Distribution of the difference between Model and Operator Picks, coloured by uncertainty method 1.

Considering Figure 4 above, we see that uncertainty tends to decrease away from 0. Recall that this metric is based on variance between the two sets of operator picks. It is therefore intuitive that where two operators don't pick consistently our model also struggles to pick accurately. Of particular note are those examples were the model has picked early. Where difference is 0 to -50, we see quite distinctly that operator variance is quite high. This likely arises as we have trained the model on the interpreter pick as opposed to the field pick.

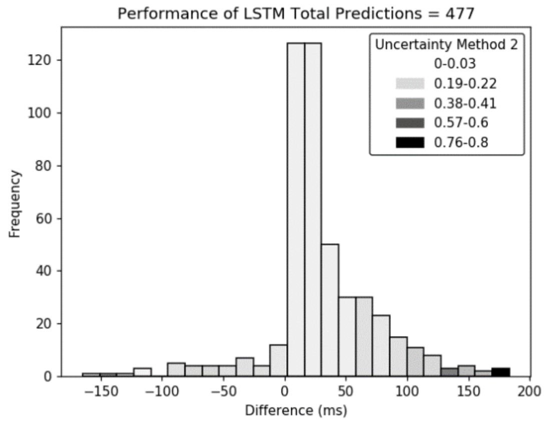

Figure 5. Distribution of the difference between Model and Operator Picks, coloured by uncertainty method 2.

With regards to the noise components of the signal, the results show, where the model contains little noise it picks consistently with the operator. As noise increases the difference between model pick and operator pick again starts to increase, with the model picking increasingly later. This finding is confirmed and exemplified by the first break clarity metric, Figure 6, where again we see that lower uncertainty is associated with better model performance, with late picks corresponding with higher uncertainty values.

In addition, the picks made by the ML model are not as consistent as those made by the operators. This is visualised in Figure 7, showing how comparatively dense the distribution is when comparing operator and model performance. We expect the model to perform worse when compared to the operator as operator picks are ultimately the metric, we use to evaluate model performance. This study however shows the viability of LSTM to accurately pick refraction data.

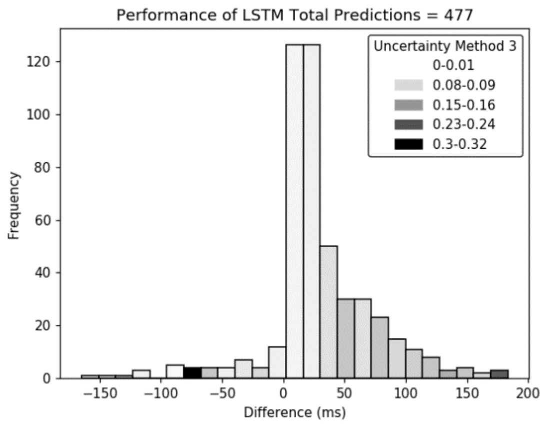

Figure 6. Distribution of the difference between Model and Operator Picks, coloured by uncertainty method 3.

CONCLUSIONS

From the results of this preliminary study we can see that the methodology used holds some promise. Overall an LSTM model trained on operator picked data is able to replicate these picks in a majority of cases. Noise and first break clarity clearly limit the ability of representation learning to accurately replicate operator picks.

In future work I evaluate alternatives to representation learning using feature space engineering, which improve upon the results of this study. It also integrates a larger representation of the geological context and site-specific challenges, a deeper evaluation of the results considering the geophysical theory involved, and a larger dataset.

It would also be interesting to train two models; one on field-based picks and one on interpreter picks. If results could be replicated accurate this could represent a way to capture a given operators skill and retain it for future training and evaluation.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisory team at the University of Western Australia; Prof. Eun-Jung Holden, Dr Daniel Wedge, and Dr Tom Horrocks for their feedback throughout these early stages of my project. I'd also like to extend my thanks to the team at Rio Tinto Resource Evaluation; Tom Green, Carlie Byrne, and Matthew Lawrence, as well as the team at Rio Tinto Exploration who were involved with the original acquisition.

REFERENCES

Asim, K. M., Idris, A., Iqbal, T. and Martínez-Álvarez, F. (2018) 'Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification', Soil Dynamics and Earthquake Engineering. Elsevier Ltd, 111(April), pp. 1–7. doi: 10.1016/j.soildyn.2018.04.020.

Boschetti, F., Dentith, M. D. and List, R. D. (1996) 'A fractal-based algorithm for detecting first arrivals on seismic traces', Geophysics, 61(4), pp. 1095–1102. doi: 10.1190/1.1444030.

Chu, C. K. P. and Mendel, J. M. (1994) 'First Break Refraction Event Picking Using Fuzzy Logic Systems', IEEE Transactions on Fuzzy Systems, 2(4), pp. 255–266. doi: 10.1109/91.324805.

Coppens, F. (1985) 'First arrival picking on common-offset trace collections for automatic estimation of static corrections', Geophysical Prospecting, 33, pp. 1212–1231. doi: 10.1111/j.1365-2478.1985.tb01360.x.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R. and Schmidhuber, J. (2017) 'LSTM: A Search Space Odyssey', IEEE Transactions on neural networks and learning systems, 28(10), pp. 2222–2232. doi: 10.1109/TNNLS.2016.2582924.

Haederle, M., Mathews, L. and Enright, M. (2016) 'Development and implementation of the sparse refraction method to exploration for detrital Fe deposits', ASEG-PESA-AIG 25th Geophysical Conference and Exhibition, (Figure 1), pp. 702–708.

Hatherly, P. J. (1982) 'A computer method for determining seismic first arribal times', Geophysics, 47(10), pp. 1431–1436.

Hawkins, L. V. (1961) 'The Reciprocal Method of Routine Shallow Seismic Refraction Investigations', Geophysics, 26(6), pp. 806–819.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', Neural Computation, 9(8). doi: 10.3109/21695717.2013.794593.

Johnson, Stephen, C. (1967) 'Hierarchical Clustering Schemes', Psychometrika, 32(3).

McCormack, M. D., Zaucha, D. E. and Dushek, D. W. (1993) 'First-break refraction event picking and seismic data trace editing using neural networks', Geophysics, 58(1), pp. 67–78. doi: 10.1190/1.1443352.

Palmer, D. (1981) 'An introduction to the generalized reciprocal method of seismic refraction interpretation', Geophysics, 46(22), pp. 1508–1518. doi: 10.1088/0957-4484/23/45/455702.

Pealdi, R. and Clement, A. (1972) 'Digital Processing of Refraction Data-Study of First Arrivals', Geophysical Prospecting, 20(03), pp. 529–548. Available at: http://www.earthdoc.org/publication/publicationdetails/?publication=34282.

Robinson, A. J. and Fallside, F. (1987) 'The Utility Driven Dynamic Error Propagation', p. 28.
Sabbione, J. I. and Velis, D. (2010) 'Automatic first-breaks picking: New strategies and algorithms', Geophysics, 75(4), pp. V67–V76. doi: 10.1190/1.3463703.

Salvador, S. and Chan, P. (2018) 'FastDTW: Toward accurate dynamic time warping in linear time and space', Intelligent Data Analysis, 11(5), pp. 561–580. doi: 10.3233/ida-2007-11508.

Samé, A., Chamroukhi, F., Govaert, G. and Aknin, P. (2011) 'Model-based clustering and segmentation of time series with changes in regime', Advances in Data Analysis and Classification, 5(4), pp. 301–321. doi: 10.1007/s11634-011-0096-5.

Wang, Q., Guo, Y., Yu, L. and Li, P. (2017) 'Earthquake Prediction based on Spatio-Temporal Data Mining: An LSTM Network Approach', IEEE Transactions on Emerging Topics in Computing, 6750(c), pp. 1–1. doi: 10.1109/TETC.2017.2699169.

Wannes Meert, Craenendonck, T. V. (2018) 'Wannesm/Dtaidistance V1.1.2.' doi: https://doi.org/10.5281/zenodo.1314205.

Williams, R. and Zipser, D. (1992) 'Gradient-based learning algorithms for recurrent networks and their computational complexity', Back-propagation: Theory, Architectures and Applications. doi: 10.1080/02673039508720837.