Preliminary Results of Automatic P-Wave Regional Earthquake Arrival Time Picking Using Machine Learning with STA/LTA As the Input Parameters

Y H Lumban Gaol¹, R K Lobo¹, S S Angkasa¹, A Abdullah¹, I Madrinovella¹, S Widyanti¹, A Priyono¹,², S K Suardja¹, A D Nugraha², Z Zulfakriza², S Widlyantoro³,⁴, M Luqman Hakim⁴, K H Palgunadi⁵

¹Faculty of Exploration and Production Technology, University of Pertamina, Jl. Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta 12220, Indonesia
²Global Geophysics Research Group, Faculty of Mining and Petroleum Engineering Institute of Technology Bandung, Jalan Ganesa No. 10, Bandung 40132, Indonesia
³Faculty of Engineering, Maranatha Christian University Bandung, Jalan Surya Sumantri No 65, Bandung 40164, Indonesia
⁴Agency for Meteorology, Climatology and Geophysics (BMKG), Jalan Angkasa I, No. 2, Kemayoran, Jakarta, Indonesia
⁵Physical Science and Engineering, King Abdullah University of Science and Technology, Saudi Arabia.

*Corresponding author: s4ndy104@gmail.com

Abstract. The Short Term Averaging/Long Term Averaging (STA/LTA) has been widely used to detect earthquake arrival time. The method simply calculates the ratio of moving average of the waveform amplitude at short and long-time windows. However, although STA/LTA signals can distinguish between real events and noise, we still recognize some lack of accuracies in first P wave arrival pickings. In this study, we attempt to implement one machine learning method popularly, Artificial Neural Network (ANN) that employ input, hidden and output layer similar as human brain works. Note that in this study, we also try to add input parameters with another derivative signal attributes such as Recursive STA/LTA and Carl STA/LTA. The processing step started by collecting event waveforms from the Agency of Meteorology, Climatology and Geophysics. We chose regional events with moment magnitude higher than 3 in the Maluku region Indonesia. Next, we apply all STA/LTA attributes to the input waveforms. We also tested our STA/LTA with synthetic data and additional noise. Further step, we manually picked the arrival of P wave events and used this as the output for ANN. In total, we used 100 events for arrival data training in P wave phases. In the validation process, an accuracy of more than 0.98 can be obtained after 200 iterations. Final outputs showed, that in average, the difference between manual picking and automatic picking from ANN is 0.45 s. We are able to increase the accuracy by band pass filter (0.1 – 3 Hz) all signal and improve the mean into 0.15s difference between manual picking and ANN picks.
1. Introduction
Traditionally, the determination of P wave was done manually by an observer/operator through visual inspection. The manual picking method gives us accurate results, but it also requires a lot of time since the massive number of earthquake data. The results will also be different from each operator, depending on their experiences. [1] proposed a method called Short Term Averaging/Long Term Averaging (STA/LTA) to automize the first break picking process that also widely used to replace the manual picking method. However, this method is only effective when the earthquake has a high signal-to-noise ratio (SNR) and will be difficult to identify the first break of earthquake with low SNR. The lack of STA/LTA method in identifying the first break of earthquake with low SNR motivate us to try applying machine learning approach and STA/LTA attributes in predicting the P-wave earthquake arrival time. [2] have tried to improve the STA/LTA method's performance by using a fitness function. [3] used machine learning algorithms to predict the earthquake first-break using 11 input parameters that will be trained to learn its pattern to predict the first break of a new earthquake data. In this paper, we try to apply machine learning algorithm called Artificial Neural Network (ANN) by combining three types STA/LTA (Classic, Recursive and Carl STA/LTA) as the input parameters [3]. These parameters will be trained using Keras, a Python’s newest machine learning library. The result expected from this research is a Python-based algorithm that can do automatic first-break picking.

2. Methods
2.1. Artificial Neural Network (ANN)
ANN is one of the machine learning algorithms that work similarly to the human brain system. ANN is a collection of connected units or nodes called neurons that can be separated into three main parts (architecture): the input layer, hidden layer, and output layer (Figure 1). The general equation of ANN is [4]:

\[ y = f \left( \sum_{i=0}^{n} (x_i W_i) + B \right) \]

where \( y \) is output, \( x \) is input, \( W \) is weight and \( B \) is bias

![Figure 1. ANN general architecture](image)

2.2. Short Term Averaging/Long Term Averaging (STA/LTA)
Short Term Averaging/Long Term Averaging (STA/LTA) is a method that works similar to the moving average of two particular windows. A possible arrival time is declared when the ratio between a short-term average (STA) of the signal and its long-term average (LTA) exceeds a certain threshold parameter that calculates on a specific sliding/moving windows. Visually, we can quickly identify earthquake arrival time from the STA/LTA curve that will show a higher spike than the overall curve.

In this paper, we calculate three types of STA/LTA (Classic, Carl, and Recursive STA/LTA) methods that are provided by Obspy, a python library for seismic processing. Generally, all of these types of STA/LTA are work similar, and the only difference is their equation that is shown below: Classic STA/LTA [5]:

[5]
where \( STA \) is short term average, \( N_{STA} \) is STA window length, \( A \) is earthquake signal amplitude, \( LTA \) is long term average, \( N_{LTA} \) is LTA window length, \( RSL \) is the ratio of STA and LTA.

Recursive STA/LTA [6] [7]:

\[
STA_{(i+1)} = \frac{1}{N_{STA}} (A_{i} - STA_{i}) + STA_{i}
\]

\[
LTA_{(i+1)} = \frac{1}{N_{LTA}} (A_{i} - LTA_{i}) + LTA_{i}
\]

\[
RSL_{(i+1)} = \frac{STA_{(i+1)}}{LTA_{(i+1)}}
\]

where \( STA \) is short term average, \( N_{STA} \) is STA window length, \( A \) is earthquake signal amplitude, \( LTA \) is long term average, \( N_{LTA} \) is LTA window length, \( RSL \) is the ratio of STA and LTA

Carl STA/LTA [7]:

\[
ETA = star - (ratio \times ltar) - abs(STA_{t} - LTA_{t}) - quiet
\]

where: \( ETA \) is the detector response – a value over 0 means detection, \( STA \) is the short term moving average of signal, \( LTA \) is the long term moving average of signal, \( star \) is the short term moving average of absolute value of signal and \( LTA \) difference, \( ltar \) is the long term moving average of star, ratio, and quiet-sensitivity parameters.

2.3. Data Processing and General Workflow

In this study, we used 106 local-earthquake waveforms with sampling rate of 20 samples/second and magnitude > 3 that recorded by BMKG station from November 06, 2019 to November 15, 2019. We apply a bandpass filter with an average frequency range is 0.5 - 2.5 Hz to increase the signal-to-noise ratio of the waveforms. Next, we manually picked the first break and calculate the ratio of three types STA/LTA that will be used as input parameters. Then, we give a label to the data with 1 and 0; 1 for the first-break and 0 for not first-break. We used 11 samples of each waveform with detail as one sample with label 1 is the first-break, and ten samples with label 0 are non-first-break (five samples before and after the first break). The dataset will be normalized to simplify the calculation process. After that, we train the dataset and do blind testing to check if our model is already good enough or not.

3. Results and Discussions

We trained 106 pairs datasets that consist of kurtosis and skewness of each event. This machine learning model's accuracy has reached 91% after 200 iterations with model loss below 0.25. A good model loss is getting closer to 0 because it means that machine learning's predictive results have no difference to the actual value.
A. Synthetic Data

We tested our machine learning model on three types of synthetic data with sampling rate of 4 samples/second (Figure 2). On clean synthetic data, the computer predicted the first break precisely the same with our manual pick. Meanwhile, on medium and high noise synthetic data there are few seconds of differences to our manual pick. (1.25 second and 2.75 second [Table 1]) miss to our manual picks on medium and high noise synthetic data. It means that our model is already working well on low noise earthquake waveform and will be less accurate on more noise earthquake waveform.

![Figure 2. STA/LTA of Clean Synthetic Data](image)

Table 1. Analysis of first-break picking prediction on synthetic data

|                      | Sample position on manual first-break picking | Sample position on auto first-break picking | Differences |
|----------------------|----------------------------------------------|--------------------------------------------|-------------|
| Clean synthetic data | 588\textsuperscript{th} (147 second)          | 588\textsuperscript{th} (147 second)       | 0           |
| Medium noise synthetic data | 588\textsuperscript{th} (147 second) | 593\textsuperscript{th} (148.25 second) | 5 samples (1.25 second) |
| High noise synthetic data | 607\textsuperscript{th} (151.75 second) | 596\textsuperscript{th} (149 second) | 11 samples (2.75 second) |

B. Real Data

After that, we tried to apply our machine learning model on two real data or events recorded on GLMI station, one of BMKG stations located on Galela, North Maluku Province. Based on the results on
synthetic data, we understand that noise greatly effects the accuracy of our model to determine the first break, so we are going to try our machine learning model on raw and filtered waveform to see how good it works on real earthquake waveform. Event 1 (Figure 3) is an earthquake with a magnitude of 5.1 and located in Minahasa Peninsula, Sulawesi. Meanwhile, Event 2 (Figure 4) is an earthquake with a magnitude of 6.1 and located in Papua. From the pictures, we can see that our model is working accurately enough to detect the first break on both raw and filtered waveform. The largest error is 0.65 seconds missed to the real first break (manual pick) on Event 1 raw waveform but the error decrease on Event 1 filtered waveform with the 0.15 seconds shift to the real first break (Table 2).

Figure 3. Auto first-break picking on real data 1 (raw and filtered)

Figure 4. Auto first-break picking on real data 1 (raw and filtered)

| Event    | Differences on raw data (s) | Differences on filtered data (s) |
|----------|-----------------------------|---------------------------------|
| Event 1  | 13 samples (0.65s)          | 3 samples (0.15s)               |
| Event 2  | 5 samples (0.25s)           | 4 samples (0.2s)                |
4. Conclusion

We have successfully conducted the application of an artificial neural network in predicting the first break of earthquakes. The results show satisfying results on both synthetic and real data. The largest error or difference between our machine learning model and manual picking is 0.65 seconds in raw Event 1 waveform. From the test on synthetic and real data, we also understand that noise greatly affects the accuracy of our model to determine the first break so how effective the waveform is filtered will become one of the most important stages in order to have a satisfying result. We still can improve our machine learning model by adding more data to be trained, and we also need further validation on more data to see how far our machine learning model can work properly.

References

[1] Allen R V 1978 Automatic earthquake recognition and timing from single traces Bulletin of the Seismological Society of America 68(1) 521–1,532
[2] Maurizio V, Satriano C, Lomax A 2012 Automatic Picker Developments and Optimization: A Strategy for Improving the Performances of Automatic Phase Pickers Seismological Research Letters 83(3) 541-554
[3] Mężyk M, & Malinowski M 2019 Multi-pattern algorithm for first-break picking employing open-source machine learning libraries Journal of Applied Geophysics 170 103848
[4] Aggarwal C C 2018 Neural Networks and Deep Learning In Neural Networks and Deep Learning
[5] Chen Y 2017 Automatic microseismic event picking via unsupervised machine learning Geophysical Journal International 212(1) 88–102
[6] Sharma B K, Kumar A, Murthy V M 2010 Evaluation of Seismic Events Detection Algorithms Journal Geological Society of India 75(3) 533-538
[7] Polkowski M, Plesiewicz B, Wiszniewski J, Wilde-Piórko M, & Group P W 2016 Local Seismic Events in the Area of Poland Based on Data from the PASSEQ 2006–2008 Experiment Acta Geophysica 64(6), 2092–2113