Single-Station Earthquake Location Using Deep Neural Networks

S. Mostafa Mousavi
Department of Geophysics
Stanford University
Stanford, CA, 94305
mmousavi@stanford.edu

Gregory C. Beroza
Department of Geophysics
Stanford University
Stanford, CA, 94305
beroza@stanford.edu

Abstract

In seismology, earthquake location is commonly done based on observed arrival times at multiple stations using a velocity model for the region and through an iterative inversion process. This makes the location estimation of earthquakes that are sparsely recorded - either because they are small or because stations are widely separated - difficult. Here, we present a fast approach based on deep neural networks to directly locate earthquakes using single-station observations. We use a multi-task temporal convolutional neural network in a Bayesian framework to learn epicentral distance and P travel time from 1-minute seismograms along with their epistemic and aleatory uncertainties. We design a separate multi-input network using standard convolutional layers to estimate the back-azimuth angle, and its epistemic uncertainty. Using this information, we estimate the epicenter, origin time, and depth along with their confidence intervals.

1 Introduction

Recent years have seen a renewed surge of interest in applying machine-learning techniques to seismic signal processing and earthquake monitoring. Despite impressive achievements in developing deep-learning approaches for earthquake detection and phase picking, earthquake location remains a challenging and unsolved task.

Lomax et. al., [Lomax et al.(2019)Lomax, Michelini, and Jozinović] expanded this approach by classifying seismic waveforms into a larger number of classes including: event/noise (1 class), station–event distance (50 classes), station–event azimuth (36 classes, each 10 degrees), event magnitude (20 classes), and event depth (20 classes); however, their model did not generalize well and suffered from high error rates. On the other hand multistation approaches (e.g. [Kriegerowski et al.(2019)Kriegerowski, Petersen, Vasyura-Bathke, and Ohrnberger]) result in a better performance by learning the move out patterns for specific station configuration at a local region.

Neither these, nor other, neural networks applied to earthquake data quantify uncertainty in their output. Machine learning can be thought of as inferring plausible models that explain data and can be used to make predictions about unseen data. Uncertainty plays a key role in that process of quantifying the reliability of those predictions. Data can be consistent with different models and the question of what model is appropriate based on such data is uncertain. Predictions using future data are also uncertain [Ghahramani(2015)].

In this study, we approach single-station earthquake location as a regression problem using two separate Bayesian neural networks. For learning epicentral distance and P travel time we designed a multi-task temporal convolutional network. The network consists of causal dilated convolutions and residual connections that estimate epicentral distance and P travel time simultaneously along

AI for Earth Sciences Workshop at NeurIPS 2020.

with their epistemic and aleatory uncertainties. We use a separate multi-input network with standard convolutional layers to estimate the back-azimuth and its epistemic uncertainty. Using this information, we estimate the epicenter, origin time, and depth along with their confidence intervals. We use a global data set for building the model and for demonstrating its performance. The proposed approach can be used for rapid earthquake source characterization using a limited number of observations. This can have many different applications including in earthquake early warning systems and in locating earthquakes that are sparsely recorded.

2 Method

Uncertainty Estimation Here, we use Monte Carlo dropout sampling [Gal and Ghahramani(2016)] for variational Bayesian approximation. Dropout randomly removes network units during the training and by doing this samples from a number of trained networks with reduced width. For test data, dropout approximates the effect of averaging the predictions of thinned networks using the weights of the un-thinned network. Gal and Ghahramani [Gal and Ghahramani(2016)] showed that a neural network with with dropout applied during both training and test times is mathematically equivalent to an approximation to the probabilistic deep Gaussian process. This is equivalent to Monte Carlo sampling from an approximate posterior distribution over models that find an approximating distribution ($q(\boldsymbol{W})$) with minimum KL divergence to the posterior probability distribution. This technique is computationally efficient, and unlike other approximations, can be easily applied to large and complex networks.

Simultaneous prediction of output and uncertainty by network makes the model robust to noisy data. The loss function acts as an intelligent regression function that allows the network to learn to attenuate the effect of erroneous labels by adapting the residual's weighting [Kendall et al.(2015)Kendall, Badrinarayanan, and Cipolla].

Temporal Convolution Networks Earthquake signals contain sequential information. The temporal dependencies among different components of an earthquake signal arise from the physics of elastic wave propagation and motivate the use of neural networks that are capable of sequence modeling. Recent studies have also demonstrated that certain convolutional architectures can outperform recurrent architectures across a diverse range of sequential-modeling tasks and data sets, while demonstrating the same capacity with longer effective memory [Bai et al.(2018)Bai, Kolter, and Koltun]. These convolutional architectures have the advantages of architectural simplicity, parallelism, flexible receptive field size, stable gradients, low memory requirements for training, and variable length inputs. Such temporal convolution networks, are a family of autoregressive feed-forward models with causal dilated convolutions and residual connections.

Network Architecture We designed two separate networks, one for predicting the epicentral distance and P travel time and the other for back-azimuth estimation.

The dist-PT network is a multi-task temporal convolutional network consisting of 1D convolutional layers where convolutions are causal and dilated. The input to the network is a 6000×4 matrix where the first three rows are 3-component waveforms (each 1 minute long with 100 samples per second) and the last row is a binary vector where values between P and S arrival times are set to 1 and the rest to zero. The last vector highlights the part of the waveforms that contain the most important information for the regression tasks. This helps for a faster convergence during the training.

The main body of the dist-PT network consists of 11 dilational convolution layers (dilation rate doubles for each layer) each with a relu activation function and 20 kernels of size 6. At the end, network has two fully connected layers each with a linear activation function and two neurons. We applied dropout to every dilated convolutional layer in the network and trained the model with a dropout rate of 0.20. The aleatory uncertainties are implicitly learned from a customized loss function during the training without a need for uncertainty labels.

Back-azimuth estimation is a continuous orientation prediction, which prohibits the direct use of a typical L2 loss function because the angle is in a non-Euclidean space. To handle this problem, we represent back-azimuth angles, *baz*, as points on a unit circle $baz = (\cos\ \theta,\ \sin\ \theta)$ during the training, and convert the predicted 2D points to the back-azimuth angles during testing.

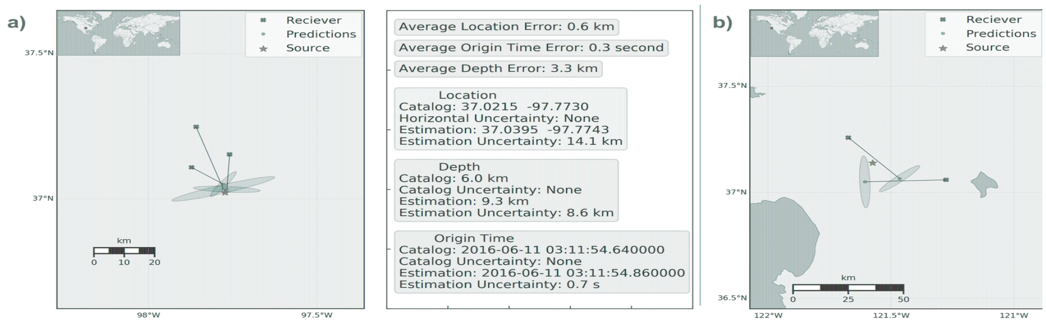

Figure 1: Performance on the test dataset: (a) individual predictions of a single event from three stations in southern Kansas; (b) an example of bad location estimation for an event in northern California.

The back-azimuth network primarily consists of 1D convolutional layers and has two inputs: 1) a (150×3) matrix (0.5 second before and 1 second after the P arrival); 2) a $(7{\times}3)$ matrix consisting of the covariance matrix, eigenvalues, and the eigenvectors derived from the 3-component waveforms for the same time window. We convolve the two input matrices with 4 and 1 convolutional layers and feed them into two fully connected layers with 100 and 2 neurons respectively to predict the coordinates of the back-azimuth angle on the unit circle. All the other layers have relu activations, except for the last fully connected layer.

3 Data

We use the STanford EArthquake Dataset (STEAD) [Mousavi et al.(2019)Mousavi, Sheng, Zhu, and Beroza] for the training and testing of the models. STEAD is a global dataset of labeled 3-component seismic waveforms (earthquake and non-earthquake). The full dataset consists of more than 1.0 million one-minute seismograms associated with 450,000 earthquakes that occurred between January 1984 and August 2018. Here, we only use earthquake waveforms recorded at epicentral distances of less than 110 km with signal-to-noise ratio of 25 decibels and higher. We only use stations for which north-south and east-west components are properly aligned to their correct geographic orientations. Based on these criteria, we randomly select 150,000 waveforms to be used for the training (% 80) and testing (% 20) of the networks. Waveforms are 1 minute in duration with a sampling rate of 100 HZ and were band-passed filtered from 1-45 HZ.

4 Results

We used the distance and back-azimuth predictions to estimate the epicenter of the associated event for each observation in the test set. We calculate the error ellipse for each epicentral location based on estimated uncertainties for distance and back-azimuth and their projections onto the reference Earth model. We use the P travel time estimates to calculate origin times and to provide a rough estimate of earthquake depth. For the depth estimation we assume that the P waves follow a straight-line path between source and station. We assumed a velocity of 5.6 km/s for the P wave and calculated the incident angle using the estimated distance the P wave has traveled together with the estimate of epicentral distance. Estimated locations and associated error ellipses for 2 events are shown in Figure 1.

We estimate locations and errors for each observation (station) and averaged for each event based on the number of available observations. Figure 1-b presents an examples with moderate errors. From the error ellipses, we can see a higher contribution of uncertainties in the back-azimuth estimations. Location estimate errors are in a reasonable range considering the reported uncertainties for the catalog locations however, we note that our location estimates are based on only single-station observations.

To get a broader view of the performance of our location estimates, we plotted the predicted epicenters paired with the cataloged locations for entire Alaska in Figure 2. We can see that the predicted

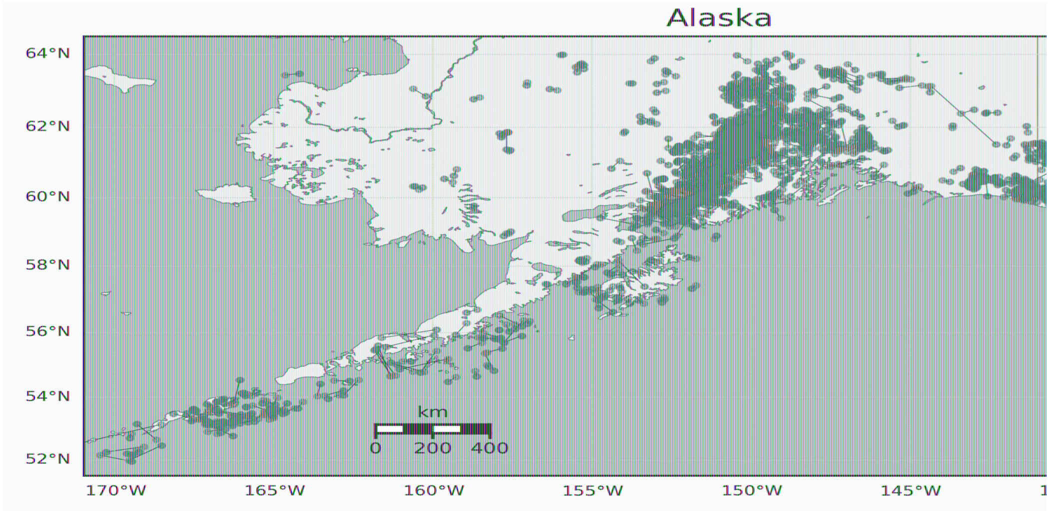

Figure 2: Single-station location predictions connected to their associated ground truth (locations in the catalogs) for Alaska

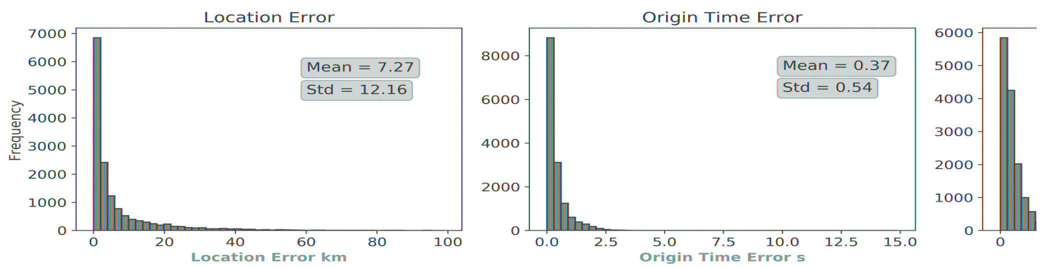

Figure 3: Statistics of location, origin time, and depth estimations for all the data (globally distributed) in the test set. Results are based on single-station estimates averaged for each event.

locations reveal the overall pattern of seismicity correctly, and that the outliers are sparse. These are averaged results for the event-based estimates without removing any high-uncertainty estimates or any weighting of the results during the averaging.

The overall performance of our locations for the entire test set can be seen from the error distributions presented in Figure 3. Our method predicts epicenter, origin time, and depth with a mean error of 7.3 km, 0.4 seconds, and 6.7 km respectively. These are in agreement with our previous observations and with regression results.

5 Conclusions

We present a successful application of deep learning for earthquake location based on single-station observations. A distinct advantage of our approach lies in its Bayesian framework, which provides an estimate of uncertainties in data and model and allows us to estimate confidence intervals in the final estimated location.

The largest uncertainty in location results is caused by uncertainties in the back-azimuth. Errors and high uncertainties in back-azimuth estimates may be due to seismic station installation and orientation error, however, learning the orientation angle using neural networks is technically challenging as well. Utilizing more advanced methods for continuous orientation estimation might improve the results. Moreover, previous studies showed the length of the window used for polarization estimation play an important role in precision of back-azimuth estimation. Optimizing this window length based on dominate signal frequency might be a potential solution for this.

The proposed method can provide a rapid estimate of earthquake location directly from single instruments. This may be useful for rapid public reporting or earthquake early warning systems.

Acknowledgements

This work was supported by the Stanford Center for Induced and Triggered Seismicity.

References

[Lomax et al.(2019)Lomax, Michelini, and Jozinović] Anthony Lomax, Alberto Michelini, and Dario Jozinović. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismological Research Letters*, 90(2A):517–529, 2019.

[Kriegerowski et al.(2019)Kriegerowski, Petersen, Vasyura-Bathke, and Ohrnberger] Marius Kriegerowski, Gesa M Petersen, Hannes Vasyura-Bathke, and Matthias Ohrnberger. A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. *Seismological Research Letters*, 90(2A):510–516, 2019.

[Ghahramani(2015)] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 2015.

[Gal and Ghahramani(2016)] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[Kendall et al.(2015)Kendall, Badrinarayanan, and Cipolla] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*, 2015.

[Bai et al.(2018)Bai, Kolter, and Koltun] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[Mousavi et al.(2019)Mousavi, Sheng, Zhu, and Beroza] S Mostafa Mousavi, Yixiao Sheng, Weiqiang Zhu, and Gregory C Beroza. Stanford earthquake dataset (stead): A global data set of seismic signals for ai. *IEEE Access*, 7:179464–179476, 2019.