Debug history

Jun ZHU

August 2, 2021

1 2021/07/08

project description:

- phase picking file path: /home/as/Data/tlPhasenet/eqphas/2014_out_dist400.txt.ok4Li
- phase picking file: # $(origin time in UTC+8 . . .) {$(station code + network code) . . . . . $(phase name) . . . $(pick time in UTC+8) . . . . . $(channel code)}
- waveform file path: /home/as/Data/tlPhasenet/xichang/20140102.002.155020.650/SC.GZA. . .
- 29497 files span from Jan 2014 to Dec 2019
- target input file format: csv
- target imput file example: fname,itp,its,channels {$(npzfile name),$(location of P-),$(location of S-),$(involved channel(s), like EHE_EHN_EHZ)}

preprocess description:

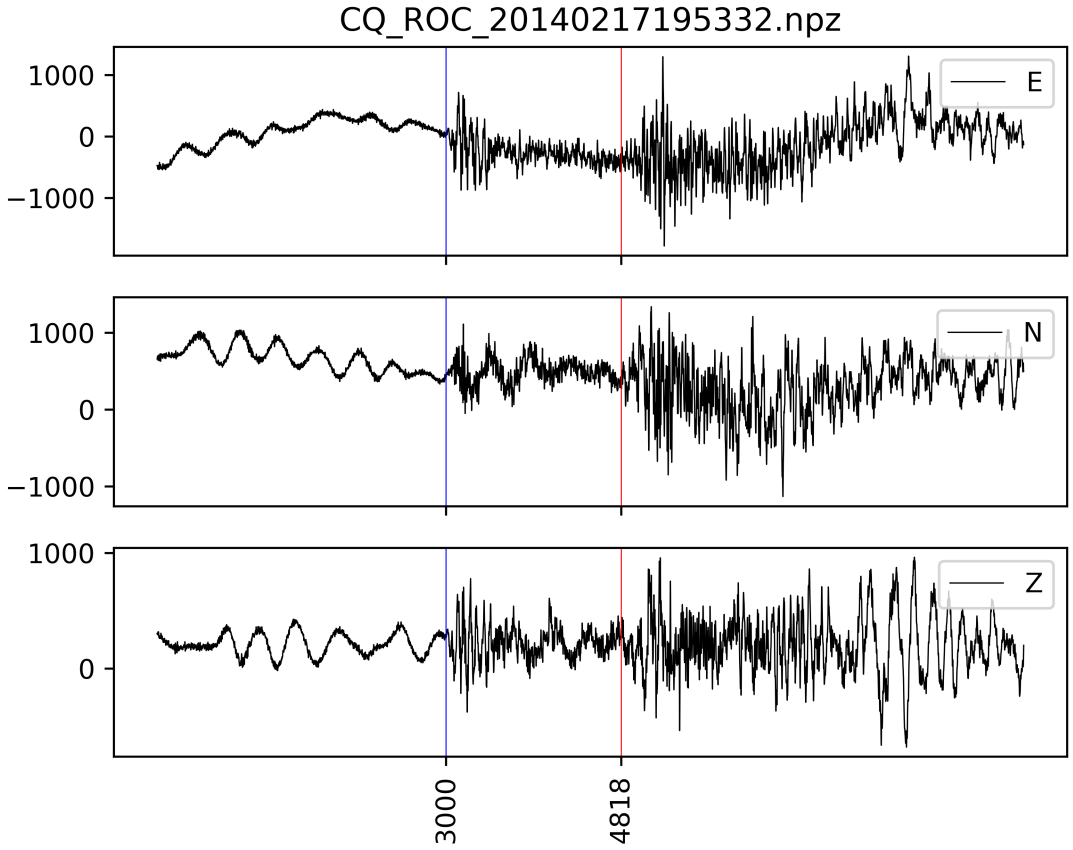
- remove empty folders inside /home/as/Data/tlPhasenet/xichang/
- only the phase picking record involving both P- and S- picks reserved
- *discard the repeating P- and S- picks*
- *some stream are not identically long in terms of each trace* (common segment of 3-components)
- *some picks repeat in the file, just conserve the first record, use the shell script named* ***removerepeat.sh***
- *some picks have over 3 traces in the folder, such as oritention code 000, 001, etc. What's worse, their sample rates are not the same. I just ignore them.*
- some picks are not involved in the 90-second training waveform segment, some are not involved in the 30-second test waveform segment. I just use the recording which involves both P- and S- picks in both train & test segment.

- screen the picks & divide the datasets based on the seismic events
- filter the waveform (the approach adopted by Chenping Chai et al.)

preprocess result overview: ~~54993~~ 52977 recordings (17919 events)

left

a sample waveform



2 2021/07/12

fix the problem of preprocessing: can't cast dtype('float32') to dtype('int32')

```
def normaliz(self, data):
   data = data - np.mean(data, axis=0, keepdims=True)
   # omit below
```

```
return
```

3 2021/07/15

evaluation metrics:

- peak probabilities above 0.5 should be counted as positive picks
- $\Delta t < 0.1$s is true positives, picks with larger residuals are counted as false positives
- mean and standard deviation are calculated on residuals ($\Delta t < 0.5s$)
- precision; recall; F1; mean; standard deviation

PhaseNet configuration:

- 779514 recordings
- different types of instruments; a wide range of SNR
- 30-s time window; 3001 points
- varied position of the arrivals within the window
- 100 Hz
- remove mean and divide by std

4 2021/07/16

Whether or not keep the recordings whose S- pick is outside the window? No.

How to freeze and load model in tensoeflow? see blog here.

5 2021/07/21

Train history of the re-train PhaseNet:

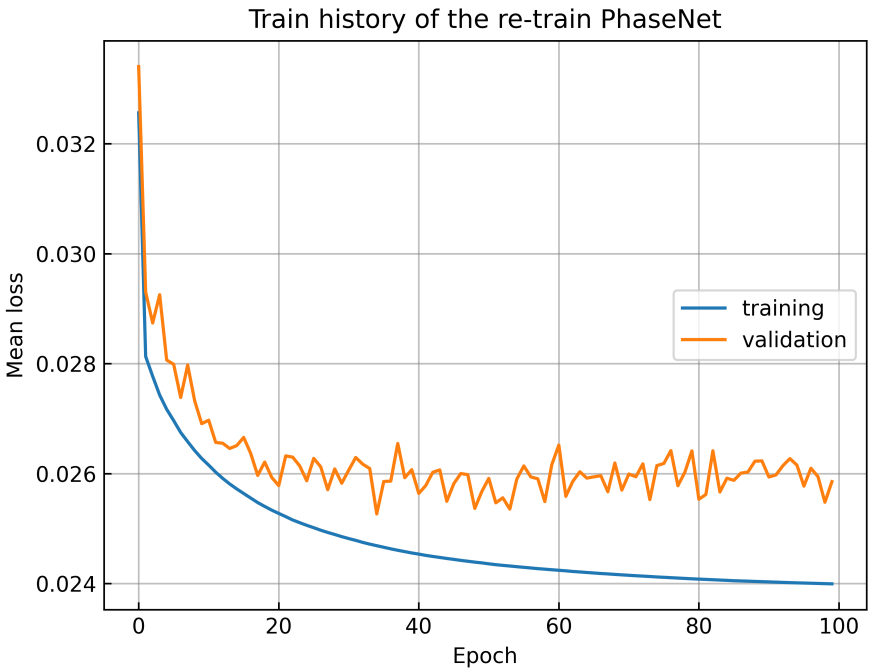


Figure 1: re-train model

Train history of the raw PhaseNet:

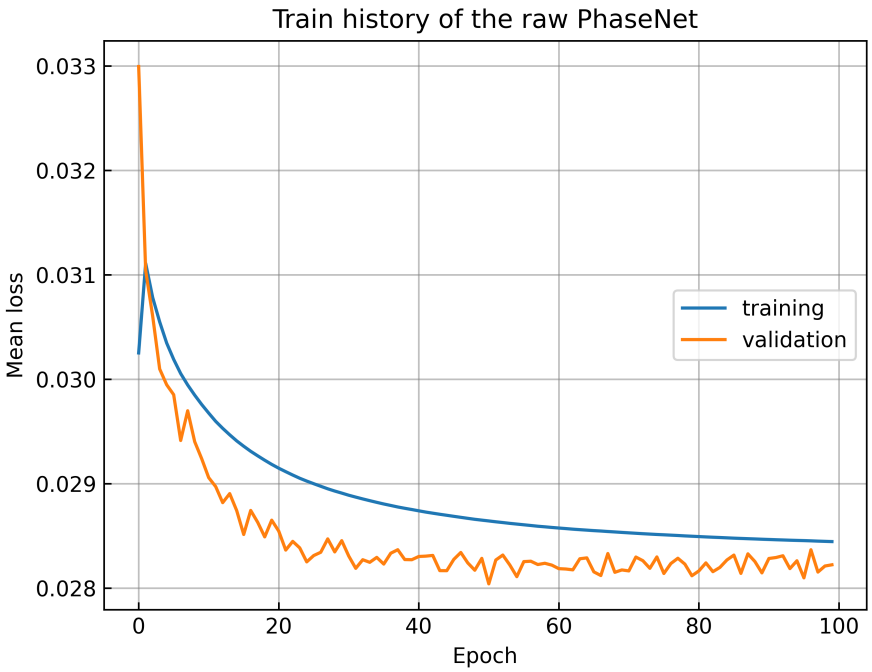


Figure 2: raw model

For pretrain model, its loss rises unexpectedly. It could be attributed to the utilization of the high learning rate. No improvement is observed even though modifying the learning rate.

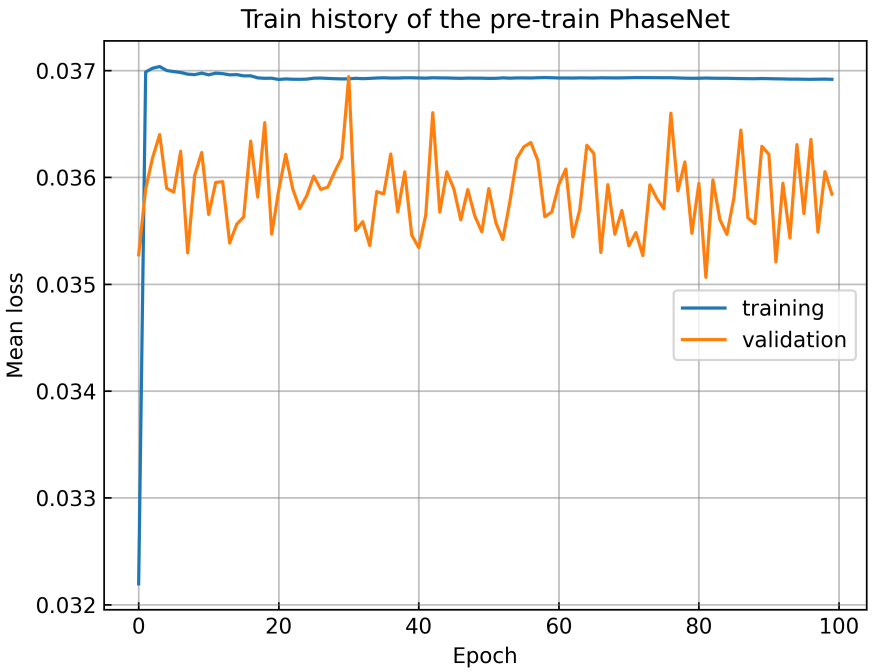


Figure 3: pretrain model

6 2021/07/29

Some problems of the pick:

- Some P- picks are mistaken as S- picks. **GZ WNT 20140502111205**

7 Reference