

Deep Learning を用いた地震波形検測

工藤 祥太、下條 賢梧 (大阪管区気象台気象防災部地震火山課)

要旨

気象庁の保有する地震波形、検測値データを用いて P 相の検測を行う Deep Learning モデルを構築した。このモデルは、現行の自動検測処理と同等の精度で P 相の検測を行うことができた。今回の条件では、学習データの個数を増加させるにつれ、精度は向上したが、10 万個以上にデータを増やしても精度の向上は見られなかった。学習済みのモデルに、周期や振幅が途中で変化するような人工的な波形を検測させたが、期待するような変化点付近の検測にはならなかった。

1. はじめに

気象庁では、関係機関から一元的に収集された地震観測データを使用し、日々震源の決定を行っている (一元化震源)。震源決定にはグループトリガ法 (清本・他、2013) や PF 法 (溜瀧・他、2016) による自動処理の結果をある程度採用するが、精度やノイズの誤検知などの問題から手動検測を行って震源計算をすることも多い。上記 2 つの自動処理では、分散比、Characteristic Function (Allen, 1978)、AR-AIC (横田・他、1981) などを用いて地震波形データの P 相・S 相を検測している。

ところで、近年、コンピュータの性能の飛躍的な向上により、コンピュータに大量のデータを与え汎用的なタスクをこなすモデルを学習させることが可能になっている。この分野は機械学習と呼ばれ発展が著しい。機械学習の中でもひととき注目を集めている Deep Learning と呼ばれるアプローチは、画像認識・音声認識・医療研究などに応用されている。

本研究では、気象庁の保有する波形データ・検測値データを組み合わせて学習データとし、Deep Learning モデルの学習を行うことによる高精度な P 相検測モデルの構築を目的とする。なお、すでに同様の研究 (Ross et al., 2018) が行われており、本研究はその手法に準拠して行う。

2. 方法

2.1 概要

P 相の検測値と対応する波形データをセットにして Deep Learning モデルに供給することで学習を進める (図 1)。波形データを入力値にしたとき、対応する検測値データに近い値が出力されるようにモデルのパラメータを更新していく。

2.2 データ

一元化震源に用いられる全観測点 (1,800 地点) の波形データのうち、2016 年 4 月 14 日~2016 年 4 月 29 日の期間に、K、k (高精度・手動) 登録の震源決定に用いられた、速度波形・鉛直成分のデータ。また、それぞれの波形に対応する P 相検測値データ。波形・検測値データセットの総数は 151,694 個。

2.3 前処理

波形データについて、オフセット・線形トレンドを除去し、バターワース特性のバンドパスフィルタ (1-20Hz) を適用。その後、P 相を中心とした 5 秒間の波形を切り出した。さら

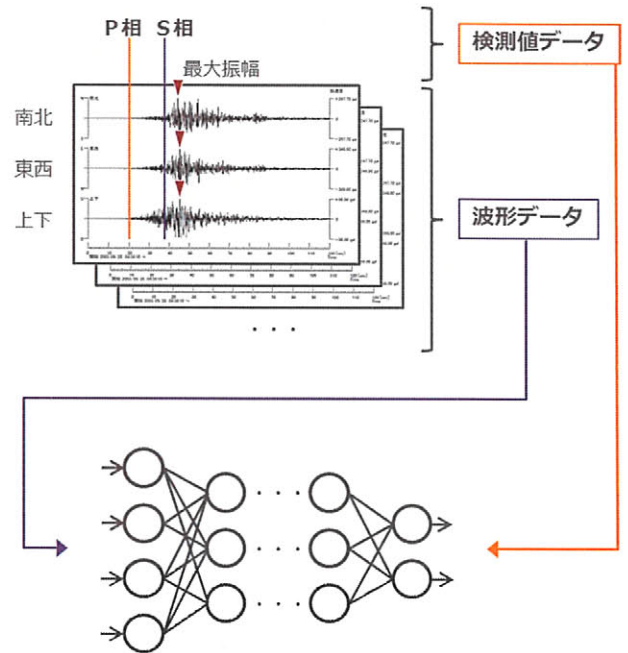


図 1 相検測タスクの学習の概要

に、訓練・検証・テストデータに 8:1:1 の割合で分配した。検証・テストデータについては、5 秒間のデータの中から連続する 4 秒間のデータをランダムに切り出し、振幅の最大絶対値で規格化した。訓練データについては、モデルへの供給時に同様の処理を行った。

2.4 モデル

波形データの中からの特徴抽出のため、1 次元畳み込み層を積み重ねた。モデルの設計を表 1、2 に示す。

表 1 モデルの設計。訓練可能なパラメータの総数: 2,943,809

層の名前	ハイパーパラメータ
A 入力層	入力サイズ: 400
B 畳み込み層 (1)	フィルタサイズ: 21 フィルタ数: 32 パディング: なし ストライド: 1 活性化関数: ReLU 出力をバッチ正規化
C プーリング層 (1)	最大値プーリング ウィンドウサイズ: 2
D 畳み込み層 (2)	フィルタサイズ: 15 フィルタ数: 64 他は B と同じ
E プーリング層 (2)	C と同じ
F 畳み込み層 (3)	フィルタサイズ: 11 フィルタ数: 128 他は B と同じ

G	プーリング層(3)	Cと同じ
H	Flatten層	1次元ベクトルに変換
I	全結合層(1)	ユニット数: 512 活性化関数: ReLU 出力をバッチ正規化
J	全結合層(2)	Iと同じ
K	全結合層(3)(出力層)	ユニット数: 1 活性化なし 出力サイズ: 1

表 2 表 1 以外のハイパーパラメータ・設定

損失関数	平均二乗誤差 (Mean Squared Error: MSE)
指標	平均絶対誤差 (Mean Absolute Error: MAE)
最適化手法	Adam
バッチサイズ	480

2.5 学習

訓練データからランダムに 480 個を選び、モデルに入力し、出力とラベル(P相検測値)から求まる損失 (MSE) が、より小さくなるようにパラメータを更新することを繰り返した。訓練データ全てを使い切る (1 エポック) ごとに MSE を訓練・検証データについて確認し、5 エポック以上検証データの MSE に減少が見られなかった場合に学習を打ち切った。ここまでの操作を 10 回行い、MSE が最も小さくなるパラメータを持つモデルを最良モデルとして選択した。

2.6 検証

実際の検測値とモデルの予測値の差について、統計量を算出した。また、学習データ数を変えて同様のことを行った。さらに、気象庁の現行システムにおける自動検測との差についての統計量も算出した。

2.7 人工波形データの検測

4 秒中のいずれかの時刻に周期・振幅が変化する sin 波を学習済みモデルに入力し、モデルの検測結果を見た。

3. 結果

学習済みモデルによる検測の例を示す (図 2)。定性的には、人間でも P 相の取り方に迷うような波形で、実際の検測値とモデルの予測値が大きく異なる傾向がある。

151,694 個のデータセットから、表 3 の通り学習用データをデータ個数を変えてランダムに取り出し、それぞれ学習を行って最良モデルを選択した。それらのモデルについて、モデルの予測値と実際の検測値との差 (残差) についての統計量を算出した (表 3)。学習データ数が増加するにつれ、残差

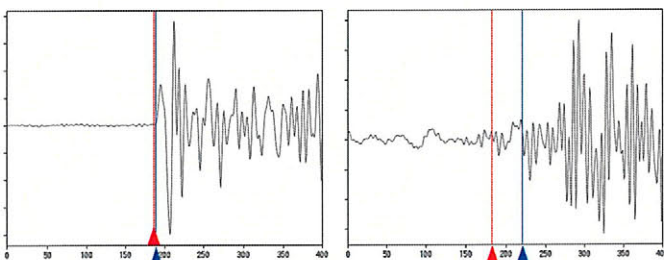


図 2 青: 実際の検測値、赤: モデルによる検測値。左は検測値がほぼ一致している例。右は比較的大きく異なる例。

の絶対平均は小さくなっているが、データ数が 100,000 以上になると 0.039 秒から変化が見られない。残差の絶対平均 0.039 秒という値は、波形データのサンプリング周波数が 100Hz (0.01 秒ごと) ということから考えて比較的小さい値である。また、最良の標準偏差も 0.069 秒とばらつきも小さい。

表 3 モデルの予測値と実際の検測値の差についての統計量

学習データ数	絶対平均[s]	標準偏差[s]
10,000	0.050	0.077
50,000	0.042	0.075
100,000	0.039	0.069
151,694	0.039	0.069

2018 年 4 月 1 日から 2018 年 4 月 30 日までの A (自動) 登録の震源決定に用いられた、波形・検測値データのセットを用いて、気象庁の現行システムによる自動検測値と、学習済みモデルによる検測値の差についての統計量を算出した。用いた震源数は 14,655 個、波形・検測値のデータセット数は 143,469 個である。学習済みモデルは学習データ数 151,694 個のものを用いた。その結果、差の絶対平均は 0.048 秒、標準偏差は 0.082 秒と比較的小さな値となった。

途中で周期や振幅が変化する sin 波を作成し、学習済みのモデルで検測させた結果を示す (図 3)。学習済みモデルが、汎用的に周期や振幅の変化点を捉えるようになっていたことを期待したが、人工的なデータについては正確な検測はできなかった。

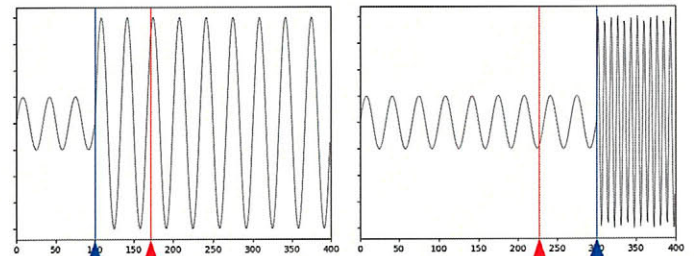


図 3 学習済みモデルによる人工的な波形の検測。青: 実際の周期・振幅の変化点、赤: モデルによる検測値。

4. 今後の展望

S 相の検測、初動方向の分類、相の分類 (P 相/S 相/ノイズの判定)、どの領域で起きた地震かの推定、遠地/近地地震の分類、現在観測している震度からの震度予測など、様々なタスクを Deep Learning を用いてこなすことを検討する。

参考文献

Z. E. Ross, M.-A. Meier, and E. Hauksson. P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120-5129, June 2018.