

機械学習モデルを用いた断層パラメータ予測に関する一検討

JEONG SANGUK^{†1} 小松一彦^{†2} 佐藤雅之^{†3} 小林広明^{†3}東北大学工学部機械知能航空工学科^{†1} 東北大学サイバーサイエンスセンター^{†2}
東北大学情報科学研究科^{†3}

1 はじめに

地震は、断層の幅や長さ、深さと傾斜角などの断層パラメータと強い関係があり、近年、リアルタイムの観測データを元にした地震の推定が行われている。しかしながら、既存研究においては、観測データと断層パラメータの関係が既に解明されている断層のデータを対象として、将来起こりうる地震の予測を行っている。そのため、既知の断層のみしか考慮することができず、新たな断層についての地震を予測することは難しい。

本研究では、機械学習を用いて、地盤変動量の観測データを特徴量として学習させることで、新たな断層における断層パラメータの予測を行う。機械学習アルゴリズムとして、従来の回帰を用いる手法とニューラルネットワーク (NN) を用いる手法の2つを検討する。回帰を用いて断層パラメータを予測する手法においては、まず、観測データからマグニチュードや観測地点から震源までの距離などを予測し、次に、それらの予測情報と観測データの両方を用いて、断層パラメータを多段階で予測する。また、NNを用いる手法においては、観測データから断層パラメータを直接予測する。NNを用いて多値の予測を行うため、予測対象となる断層パラメータ同士の相関も考慮することができる。

2 観測データと断層パラメータ

観測データは、観測地点の緯度と経度 (Lat, Lon)、東西、南北、垂直方向の地盤変動量 (DfmX, DfmY, DfmZ)、水平方向の変動と垂直方向の変動 (DfmH, DfmHZ) を合わせた7つの地盤変動パラメータである。予測対象は地震自身を発生させる断層のパラメータであり、震源の緯度と経度 (lat, lon)、深さ (depth)、走行 (strike)、傾斜角 (dip)、断層の長さ (length)、断層の幅 (width)、

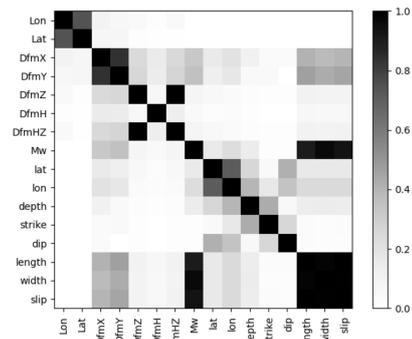


図1: 観測データと断層パラメータの相関関係。

滑り量 (slip) の8つである。

図1に観測データと断層パラメータの相関関係を表す。相関図の色が黒に近づけば近づくほど、相関関係が高いことを意味する。図1を見ると、観測地点における東西および南北方向の地盤変動量 (DfmX, DfmY) は、Mw, length, width, slip と0.5以上の相関があることが分かる。一方、depth, strike, dipは、観測データとの相関がほとんどなく、断層パラメータ同士での相関の方が強く見られる。例えば、震源地における緯度と経度 (lat, lon) と相関があることが見て取れる。これらの相関関係を活かしながら断層パラメータを予測する必要があると考えられる。

3 機械学習を用いた断層パラメータ予測

観測データから断層パラメータを予測するために、提案手法では回帰またはNNを用いると考える。回帰は単一変数の予測しか出来ないため、回帰を用いる予測においては、それぞれの断層パラメータを予測するための7つのモデルを構築する。観測データと相関の高い断層パラメータである length, width, slip は予測に向いていると考えられる。学習においては、観測データに加えて、式(1)で定義される観測地点における地盤変動量 Dfm_{uclid} も用いる。これは、観測地点における東西および南北方向の地盤変動量 (DfmX, DfmY) が断層パラメータと高い相関を持つため、地盤変動量 Dfm_{uclid}

A study on fault parameter prediction using machine learning

JEONG SANGUK^{†1}, Kazuhiko Komatsu^{†2}, Masayuki Sato^{†3}, Hiroaki Kobayashi^{†3}Department of Mechanical and Aerospace Engineering, Tohoku University^{†1}Cyberscience Center, Tohoku University^{†2}Graduate School of Information Sciences, Tohoku University^{†3}

も有用な特徴量であると考えられるためである。

$$Dfm_{uclid} = \sqrt{DfmX^2 + DfmY^2 + DfmZ^2} \quad (1)$$

回帰による方法では、2段階に分けてパラメータの予測を行う。これにより、観測データと相関の少ない depth, strike, dip の予測精度の向上を狙う。まず、観測地点における東西および南北方向の地盤変動量 ($DfmX, DfmY, DfmZ$) を用いて、式 (1) によって地盤変動量の大きさ Dfm_{uclid} を計算する。つぎに、観測データと相関の高いマグニチュード (Mw)、断層パラメータと関係性が高いと考えられる観測地点から震源までの距離 $distance1$ および緯度 0 度、経度 0 度の原点から震源までの距離 $distance2$ を回帰モデルにより予測を行う。観測データに、地盤変動量 Dfm_{uclid} , Mw , 2つの距離を特徴量として加え、8つの断層パラメータの予測を行う。

次に NN を用いる予測手法について説明する。本手法では、多値変数の予測が可能であるため、観測データを入力として7つの断層パラメータを予測する。予測対象となる断層パラメータ同士の相関も考慮することができるため、高い予測精度が期待できる。これにより、断層パラメータ同士との相関が見られる depth, strike, dip も予測できると期待できる。

4 評価

4.1 実験条件

データセットは、16748 件の地震を対象として、南海トラフ周辺の地表に置かれている 816 箇所の観測地点において測定された合計 1366 万件以上の地盤変動データである。本データセットは 7 種の 816 箇所から観測された 16748 件のデータであることから、NN を用いた予測においては、(816 × 7) のデータを 1 個として取り扱おうと学習データの数が不足し、精度低下の恐れがある。そこで (4 × 7) のデータとして扱うことで、学習データの数を十分に確保する。また、3 層の多層パーセプトロンモデルを用い、活性化関数は ReLU、損失関数は Mean Squared Error (MSE) に設定し、学習を行う。

4.2 実験結果と考察

図 2 に、回帰によって断層パラメータを予測した結果を示す。縦軸は平均絶対誤差率 (MAPE)、横軸は予測対象である断層パラメータである。青色の棒グラフは、観測データのみを用いて予測した場合、橙色は、観測データに地盤変動量 (Dfm_{uclid}) を特徴量として加え、多段階で予測した場合の結果である。

図 2 より、lat, lon, strike については低い平均絶対誤差率で予測することができていることが分かる。一方、その他の断層パラメータの平均絶対誤差率は 7.8% と 35% となった。これは、各断層パラメータにおける値のレンジが異なるため、大きな値を持つ断層パラメータにおいて、誤差率が大きくなったと考えられる。

次に、多段階で予測した結果を見ると、観測データの

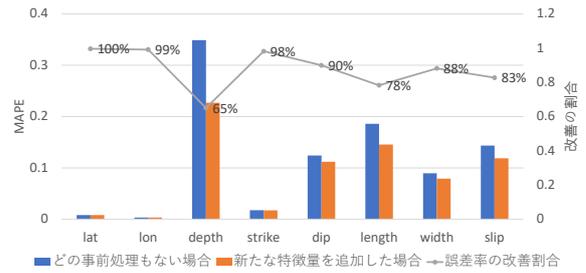


図 2: 回帰による断層パラメータの予測。

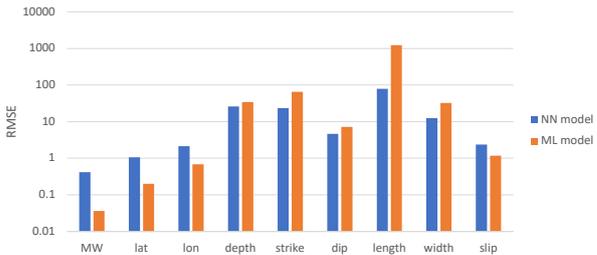


図 3: NN を用いた断層パラメータの予測。

みを用いて予測した場合よりも、誤差が小さくなっていることが分かる。これは、 Mw の予測精度が高いこと、そして Mw は他の断層パラメータと相関が高いためである。

図 3 に NN を用いた結果を示す。縦軸は二乗平均平方根誤差 (RMSE) を示す。多段階的回帰によって得られた結果を橙色、NN を用いた結果を灰色に示す。図より、NN を用いた予測は、回帰を用いた場合に比べて、strike, length などにおいては高い精度が得られていることが分かる。これは NN が予測ターゲットとである断層パラメータ間の相関を考慮できたためであると考えられる。一方、 Mw と震源の緯度・軽度など NN を用いて逆に誤差が大きなケースもあるため、これについてはより詳細な予測結果の分析が必要である。

5 おわりに

本研究では、2種類の機械学習アルゴリズムを用いて断層パラメータの予測を行った。その結果、NN を用いた予測の方が、予測ターゲットの間関係を考慮できるため、観測データと相関が少ないパラメータでも予測できることが分かった。今後の課題として、回帰・NN で得意、不得意のパラメータが存在するため、適切な使い分けについての研究が必要である。

謝辞

本研究の一部は、文部科学省「次世代領域研究開発」(高性能汎用計算機高度利用事業費補助金)「量子アニーリングアシスト型次世代スーパーコンピューティング基盤の開発」および SIP3 期サブ課題「津波災害デジタルツインの構築とスマート・レジリエンスの実現」の助成を受けて実施している。