

機械学習を用いた地震動予測において 偏ったデータセットが与える影響を軽減するための試み

Attempt to reduce the effect of biased data-set
on ground-motion prediction using machine learning

久保久彦^{*1}
Hisahiko Kubo

功刀卓^{*1}
Takashi Kunugi

鈴木進吾^{*1}
Shingo Suzuki

鈴木亘^{*1}
Wataru Suzuki

青井真^{*1}
Shin Aoi

^{*1} 防災科学技術研究所

National Research Institute for Earth Science and Disaster Resilience

Previous study [Kubo 2018] tried to construct a predictor of ground-motion index using a random forest method and strong-motion data recorded in Japan. However, the data-set is very biased and there are few strong ground-motion records. This causes the underestimation of the predictor for strong ground-motions. To overcome this problem, in this study, we suggest two approaches: one is the weighting of train data, and the other is the hybrid method integrating the conventional ground motion prediction equation and a machine learning approach. The verification using test data indicates that the hybrid method can largely improve the underestimation, although the underestimation still remains in predicting very strong ground-motions (>1000 gal).

1. はじめに

これまで著者らは機械学習を用いて地震動予測式の構築を試みてきた[久保 2018]. 地震動予測式は地震による強い揺れ(強震動)を過去の記録に基づいて経験的に予測する手法の一つであり[日本建築学会 2005], 距離減衰式とも呼ばれる. 先行研究[久保 2018]では地表最大加速度(PGA)を予測対象とし, 震央距離・モーメントマグニチュード・深さ・予測地点の地下深さ 30m までの平均 S 波速度(V_{s30}), 予測地点の地下において S 波速度が 1400m/s に達するまで深さ(D1400)の 5 つを説明変数とする予測器をランダムフォレストによって構築した. そして実記録を用いた検証から, 構築された予測器によって実用レベルの予測ができることを示した.

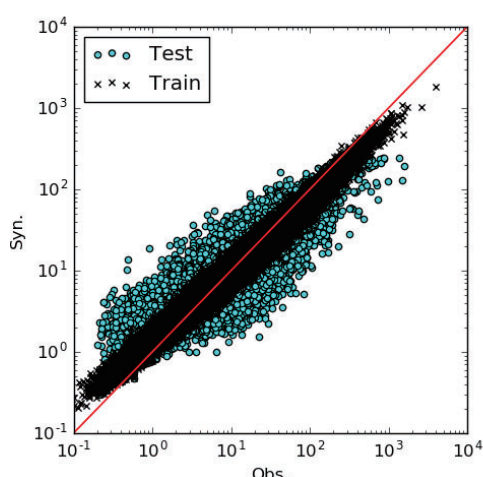


図1 [久保 2018]における観測値と予測値の比較(元は図4). ×が訓練記録を, ○がテスト記録を示す. 単位は横軸・縦軸ともに gal.

しかしながら問題も見られた. 図1に[久保 2018]の図4を示す. この図は訓練データとテストデータについて, 観測値と予測値の比較を示しており, 観測と予測が完全に一致すれば対角線上に位置することとなる. 全体的には対角線上にのっているが, 観測値が大きいほど過小評価している傾向が見られている. 特に 1000gal を超えるような非常に強い地震動を予測する際に大幅に過小な予測となっている.

図2に[久保 2018]で用いた訓練データの PGA のヒストグラムを示す. その訓練データは, 1997年4月から2015年12月までに発生した地震計 2177 イベントの際に, 防災科学技術研究所の陸海統合地震津波火山観測網 MOWLAS の全国強震観測網 K-NET・基盤強震観測網 KiK-net [功刀 2009] [Aoi 2011] の地表強震計で記録された強震記録計 288,644 記録である. 図2から 100gal 以下の比較的弱い地震動の記録が大半を占めていることがわかる. また 1000gal を超えるような強い地震動の記録は相対的に少なく, 20 個程度しかないこともわかる. このように[久保 2018]のデータセットには大きな偏りが存在しており, これが原因となり比較的弱い地震動とのフィッティングが重視された予測器が作成され, 図1で見られた強震動時の過小評価

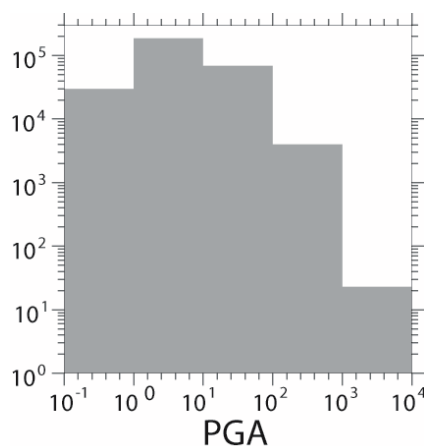


図2 [久保 2018]で用いた訓練データにおける PGA のヒストグラム

連絡先: 久保久彦, 防災科学技術研究所, 茨城県つくば天王台 3-1, 029-863-7304, hkubo@boai.go.jp

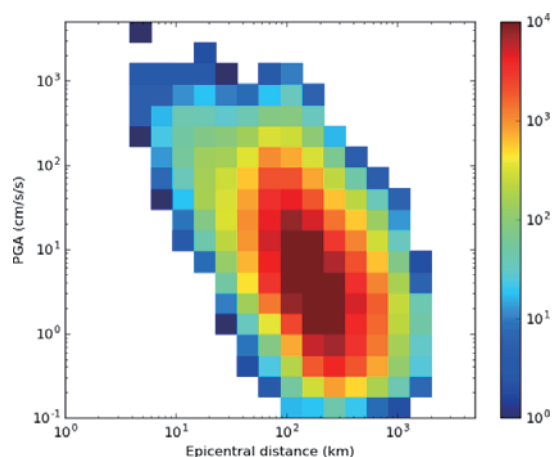


図 3 [久保 2018]で用いた訓練データにおける PGA と震央距離のヒートマップ

を引き起こしたと考えられる。

次に図 3 で[久保 2018]で用いた訓練データにおける PGA と震央距離のヒートマップを示す。PGA と震央距離の間には逆相関の関係が見られ、震央距離が小さいほど PGA が大きくなることからわかる。図 3 から分かるように、震源からの距離は地震動の強さを決める大きな要因であり、強震動を正しく予測するには、この関係を適切に評価する必要がある。しかし前述のように強い地震動がデータセット内にあまりなかったため、[久保 2018]では適切に学習できなかったと考えられる。

本研究では上記の過小評価の解決を目指して、二つのアプローチを試みる。まず地震動の強さに応じたデータセットの重みづけの使用を試みる。さらに、既往の地震動予測式による予測とランダムフォレスト予測器を組み合わせたハイブリッド手法も提案する。

2. データの重みづけを用いた場合

まずデータへの重みづけを試す。既往研究では震央距離ごとにデータをカテゴリー分けし、重みづけを行うことが多い[司 1999; Morikawa 2013]。しかしながら地震動予測の対象は強震動指標であり、それに着目してデータセットの偏りを考える方が自然であるため、ここでは PGA に着目してデータセットの重みづけを行う。

重みづけに際しては、訓練データを①1gal 以上・10gal 未満、②10gal 以上・100gal 未満、③100gal 以上・1000gal 未満、④1000gal 以上の 4 つのカテゴリーに分けた。その際 1gal 未満の記録はデータセットから取り除いている。各カテゴリーのデータ数は①が 113,829 個、②が 60,796 個、③が 3895 個、④が 23 個である、そしてカテゴリーごとに重みづけを変えた訓練データから予測器を生成し、得られた予測器の性能をテストデータを用いて検証した。データの重みづけは少ない個数のカテゴリーの重みが大きくなるよう、①は 1 倍で固定した上で、②は 1・2 倍の 2 ケース、③は 1・2・4・8・16 倍の 5 ケース、④は 1・2・4・8・16・32・64・128・256・512・1024 倍の 11 ケースをそれぞれ試した、なお訓練データは前述のとおりであるが、テストデータは 2016 年 1 月から 10 月までに発生した地震 計 223 イベントによる強震記録 31,273 記録である。データセット以外のやり方は[久保 2018]に則った。

なお、地震動記録がトリガー収録であることが機械学習の予測器による予測結果に影響するかどうかの確認も同時に行った。

震央距離 200km 以下の記録のみで構成されるデータセットを用いてこの後の解析を進めるが、結論的には特に影響は見られなかった。

データの重みづけを行った場合の解析結果例として、ここでは訓練データの重みを①が 1 倍、②が 2 倍、③が 16 倍、④が 1024 倍としたときの結果を示す。図 4 に重みづけをしなかったデータを用いた場合の結果、図 5 に重みづけを行ったデータを用いた場合の結果を示す。これらの図の比較からわかるように、データの重みづけを行うと訓練データでは問題の過小評価が改善されているが、テストデータでは改善されておらず、問題が解決できていないことがわかる。今回試したどの重みづけの場合でも問題の過小評価がテストデータでは改善されなかった。なお強い地震動の記録の重みを大きくすると、訓練データの交差検定における R^2 の平均は良化し、またランダムフォレストで得られる各説明変数の重要度において震央距離の重要度が高くなることも確認された。

データの重みづけを行っても問題の過小評価がテストデータでは改善できなかった原因としては以下が考えられる。重みづけした訓練データを学習しても、地震動の強さに大きな影響を与える震央距離(もしくは震源距離)が適切に評価することができておらず、他の説明変数によって訓練データを説明してしまい、いわゆる過学習の状態に陥っていたと考えられる。そのため問

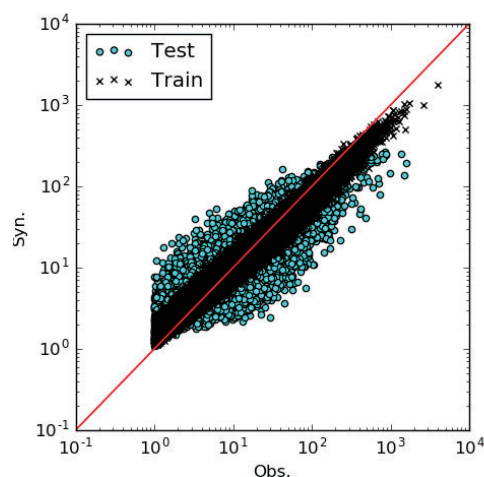


図 4 データの重みづけを適用しなかった場合のランダムフォレスト予測器による予測値と観測値の比較。

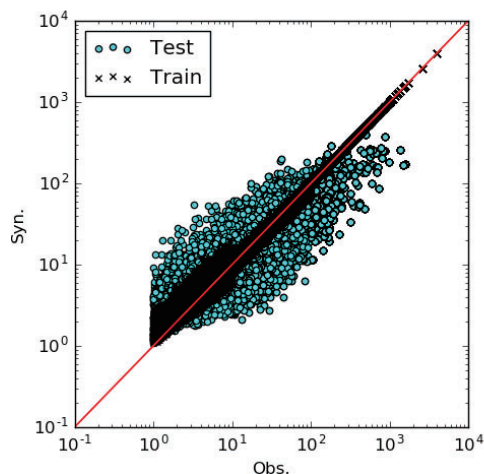


図 5 データの重みづけを適用した場合のランダムフォレスト予測器による予測値と観測値の比較。

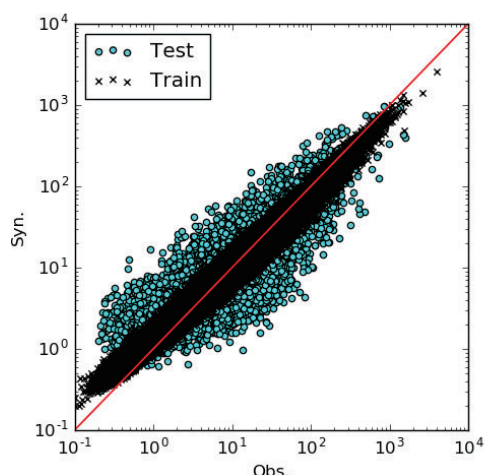


図 6 ハイブリッド法を適用した場合に得られた予測器による予測値と観測値の比較。

題の過小評価がテストデータでは解消できなかったと考えられる。なお震央距離で重みづけしてもここで述べた結果は変わらなかったことも確認している。

3. ハイブリッド手法を用いた場合

次に機械学習と既往の地震動予測式を組み合わせたハイブリッド手法を試みる。既往の地震動予測式は基礎式としてあらかじめ式を仮定した上で回帰によって式のパラメータを決めていることから、物理背景を関数形の形で考慮することができ、外挿に強いというメリットがある一方で、その関数形で表現できないことは予測できない。他方でランダムフォレストをはじめとするノンパラメトリックな機械学習手法は柔軟性が高い一方で、外挿部分、もしくはデータが少ないところでは心もとないというデメリットを持つ。そこで両者の強みを生かすために、①地震動予測式による基礎的な予測と②ランダムフォレストによる付加的な予測を足し合わせたものを予測とするハイブリッド手法を提案する。①には既往の地震動予測式 [Morikawa 2013] の model 1 を用いた：

$$\log pre = a[\min(M_w, M_{w0}) - M_w']^2 + bX + c - \log(X + d \cdot 10^{0.5 M_w}) \quad \#$$

式中の係数は地殻内地震のものを利用した。この場合のインプットは震源距離 X と M_w となる。震源距離は震央距離と深さから算出した。②には震央距離・ M_w ・深さ・ V_{s30} ・ $D1400$ を説明変数とするランダムフォレスト予測器を用いる。この予測器を作成するには観測記録から①の予測値を引いた残差を学習データとして用いる。学習データ以外のやり方は [久保 2018] に則った。

得られた結果を図 6 に示す。訓練データだけでなく、テストデータでも問題の過小評価が改善していることが分かる。他方で、訓練データの交差検定における R^2 の平均が若干悪化しており、それはテストデータでも同様である。

4. 結論

強い地震動記録が極端に少ないという学習データの偏りは、機械学習を用いた地震動予測に大きく影響を与え、強震動時の過小評価を引き起こす。本研究ではこの問題の解決法を模索し、既往の距離減衰式とランダムフォレストの両方を組み合わ

せたハイブリッド手法を用いることで強震動時の過小評価が大きく改善されることを示した。ハイブリッド手法の良い点は地震動予測式をその予測の中で用いることで、地震と震源距離の間の関係を物理背景に考慮した形で予測の中に組み込むことができる点である。ただし観測が 1000gal を超えるような場合には、改善はみられるが、依然として過小評価気味であり、更なる検討が必要である。また既往研究でよく用いられてきたデータの重みづけを行うと、機械学習を適用した場合、過学習を引き起こす可能性も示された。

参考文献

- [Aoi 2011] Aoi, S., Kunugi, T., Nakamura, H., and Fujiwara, H.: Deployment of new strong motion seismographs of K-NET and KiK-net, in Earthquake Data in Engineering Seismology, Geotech. Geol. Earthquake Eng., Vol.14, pp.167–186, Springer, Dordrecht, Netherlands (2011)
- [刃刀 2009] 刃刀卓, 青井真, 藤原広行: 強震観測—歴史と展望—, 地震 2, S19-S24 (2009)
- [久保 2018] 久保久彦, 刃刀卓, 鈴木進吾, 鈴木亘, 青井真: 人工知能学会 2018, 4Pin1-35 (2018)
- [Morikawa 2013] Morikawa, N. and Fujiwara, H.: A New Ground Motion Prediction Equation for Japan Applicable up to M9 Mega-Earthquake, Journal of Disaster Research, Vol.8, No.5, pp.878–888 (2013)
- [日本建築学会 2005] 日本建築学会: 地盤震動—(日本建築学会・編集), 丸善, 408pp (2005).
- [司 1999] 司宏俊, 翠川三郎: 断層タイプ及び地盤条件を考慮した最大加速度・最大速度の距離減衰式, 日本建築学会構造系論文集, No. 523, pp. 63–70 (1999)