

海底水圧計データと機械学習を用いた津波浸水予測における観測点の寄与度の評価

課題責任者

馬場 俊孝 徳島大学大学院社会産業理工学研究部

著者

上谷 政人*

*徳島大学大学院創成科学研究科

キーワード：津波, 津波浸水予測, 機械学習, クラスタ解析, 多層パーセプトロン

1. はじめに

機械学習の手法を適用して津波を予測する取り組みが行われるようになってきた。我々はこれまでに南海トラフの海底水圧計データを入力として、津波の浸水を多層パーセプトロンで予測する枠組みを提案した(Kamiya et al., 2022)が、観測網が整備されていない海域(高知県沖～日向灘)で大すべり域や超大すべり域が想定されるシナリオの予測精度に課題があった。また、従来の研究では各観測点の予測モデルへの寄与度には着目しなかった。

一般に機械学習モデルを構築する際の特徴量を選択することで、モデルの解釈性および汎用性の向上、学習時間の短縮が期待できる。さらに、各観測点の予測モデルへの寄与度を評価することは特徴量選択だけでなく、津波観測計を新たに整備する際の指標としての観点からも重要である。本研究では南海トラフ巨大地震の津波を対象に観測網の空白域に整備されている N-net 観測点を考慮した上で、多層パーセプトロンと SHAP という手法を用いて観測点の違いによる予測精度の違いと各観測点の寄与度を評価する。

2. 解析手法

解析対象地域は徳島県阿南市とした(図1)。はじめに、先行研究が提案した 3480 ケースの断層シナリオ(藤原他, 2020) のすべてについて、地球シミュレータを用いて高分解能津波計算を実施した。予測処理の負荷を軽減するため、k-means 法を用いて浸水深がほぼ同じ領域を予めグループ化することにより予測点数を削減した。DONET 観測点 51 地点及び DONET・東南海の観測点 55 地点と現在構築中の 36 地点の計 91 地点の計算された水圧変動データを説明変数、k-means 法によって分けられた領域内の浸水深の平均値を目的変数として、機械学習により津波浸水予測モデルを 2 パターン構築した。機械学習モデルは MLP を利用した。予測モデルの構築と最適化には Python のライブラリ PyCaret を使用した。予測モデルの構築条件は中間層を 2 層とし、中間層のノード数は 1 層目を 100、2 層目を 50 とした。中間層の活性化関数には ReLU 活性化関数を使用した。最適化アルゴリズムには Adam を使用した。過学習対策のため、中間層に L2 正則化項を考慮した。また、正則化項のハイパーパラメータ値の決定にはクロスバリデーション法を使用し、その値は $\alpha=0.7$ となっ

た。

SHAP (SHapley Additive exPlanations)は協力ゲーム理論のシャープレイ値を機械学習に応用した手法である。SHAP を用いることにより、予測モデルの予測結果に対するそれぞれの特徴量の重要度を求めることができる。これにより、ブラックボックスになりがちな機械学習モデルに対して解釈性や説明性を付与することが可能となる。

予測モデルの検証のため、内閣府(2012)が提案した 11 個の M9 クラスの地震シナリオを使って、フォワード津波計算による真値と機械学習モデルの予測値を比較した。更に、SHAP を用いてテストデータの予測結果に対する各観測点の寄与度を評価した。

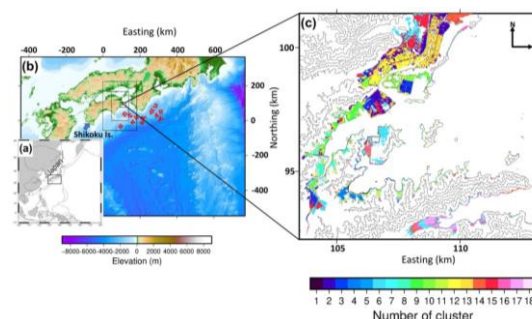


図1. (a) 研究対象地域. (b) 津波計算領域. 黒い矩形領域はネスティングレイヤーを示す. 赤の三角形は DONET 観測所の海底圧力計の位置で、座標系は日本の平面直角座標系 IV である. (c) 津波予測領域. 色は k-means 法による津波の浸水深のクラスタ化された領域を意味する. 等高線は標高を表し、間隔は 50m である.

3. テストデータの予測結果

テストデータ 11 シナリオに対する 2 種類の予測モデルの予測とフォワード計算により求めた真値を比較した散布図を図 2 に示す。従来の 51 観測点による予測モデルの決定係数 R^2 は 0.85 で、新しい観測点を加えた 91 観測点モデルの R^2 は 0.88 であり、予測精度の向上が見られた。図 3 にテストデータのシナリオ別の予測値と真値の RMSE を示す。観測網の空白域に該当するシナリオ 5 の RMSE は 51 観測点モデルが約 1.28m, 91 観測点モデルが約 0.58m

であり、予測精度の向上が見られた。しかし、紀伊水道沖に大すべり域や超大すべり域が想定されるシナリオ 3 の RMSE は 51 観測点モデルが約 0.57m, 91 観測点モデルが約 1.49m であり、予測精度の低下が見られた。

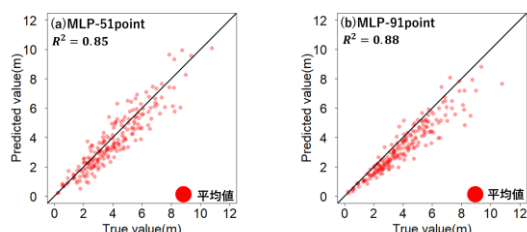


図 2. (a)51 観測点モデル, (b)91 観測点モデルによる真値（フォワード計算）と予測値との散布図。赤丸はクラスター化した領域の浸水深の平均値を示す。

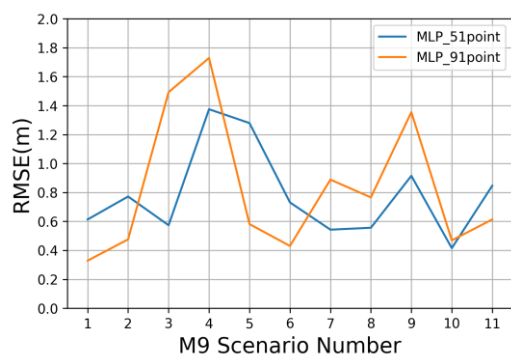


図 3. M9 シナリオの No.1~No.11 の真値（フォワード計算）と予測値の RMSE を折れ線グラフ。

シナリオ 5 のフォワード計算による津波浸水深分布と 2 種類の前測モデルによる津波浸水深分布を図 4 に示す。91 観測点モデルではリアス式海岸の地域や北部の市街地の再現性が向上していることが分かる。

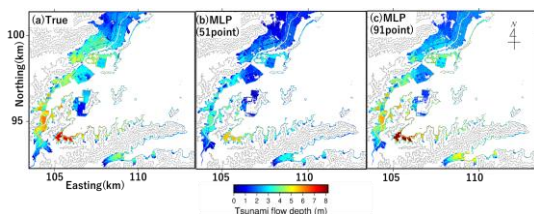


図 4. (a)M9 シナリオ 5 からフォワード計算した津波浸水深分布, (b) (c) はそれぞれ 51 観測点モデル, 91 観測点モデルで予測した津波浸水深分布を示した空間図である。等高線は標高を表し、間隔は 50m である。

4. 考察

新しい観測点を追加した予測モデルでは課題であった

観測網の空白域で大すべり域や超大すべり域があるシナリオ 5 の予測精度が向上したが、従来の観測点を用いた予測モデルで上手く予測出来ていたシナリオ 3 の予測精度が低下した。この原因を探るために 2 種類の前測モデルによるテストデータ 11 シナリオの予測結果に対する SHAP 値を求めた(図 5)。51 観測点モデルでは紀伊水道沖の観測点の寄与度が高い。一方で、91 観測点では紀伊水道沖の観測点が低下し、新たな観測網の寄与度が高まったことが分かった。これらのことから、新しい予測モデルとして次の様に解釈できる。新たな観測点を追加することで、観測網の空白地帯に大すべりが想定されている地震を上手く予測できるようになった。しかし、新たな観測点の寄与度の増加の影響により、紀伊水道沖の観測点の寄与度の低下により、シナリオ 3 の予測精度が低下したと考えられる。

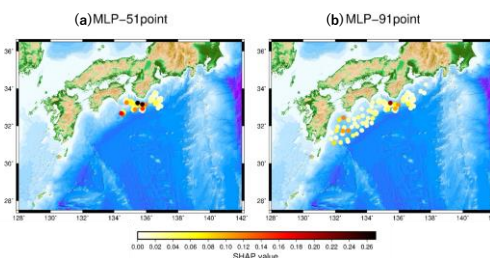


図 5. (a)51 観測点モデル, (b)91 観測点モデルのテストデータ 11 シナリオの予測結果に対する SHAP 値。

謝辞

本研究の大規模津波計算には地球シミュレータを利用させていただきました。記して、感謝いたします。

文献

- [1] 藤原広行ほか, 南海トラフ沿いの地震に対する確率的津波ハザード評価 -第一部 本編-, 防災科学技術研究所資料, 第 439 号(第一部) 71pp, https://dil-opac.bosai.go.jp/publication/nied_tech_note/pdf/n439_01m_1.pdf, 2020.
- [2] Kamiya M, Igarashi Y, Okada M, Baba T. Numerical experiments on tsunami flow depth prediction for clustered areas using regression and machine learning models, Earth Planet Space, 74, 127, 2022.
- [3] 内閣府, 2012 年南海トラフの巨大地震モデル検討会, <http://www.bousai.go.jp/jishin/nankai/model/index.html>, 最終アクセス日時 2022 年 1 月 25 日