

Evaluation of contribution of observation points to tsunami inundation prediction using machine learning

*Masato Kamiya¹, Toshitaka Baba²

1. Graduate School of Sciences and Technology for Innovation, Tokushima University, 2. Graduate School of Technology, Industrial and Social Sciences, Tokushima University

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が、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に利用する特徴量を制限することで、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。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の選択だけでなく、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Light Gradient Boosting Machine (LightGBM) と SHapley Additive exPlanations (SHAP) という手法を用いて各観測点の寄与度を評価した。LightGBMは決定木アルゴリズムに基づいた勾配ブースティングの機械学習フレームワークである。LightGBMではアンサンブル学習の手法の一つであるブースティングが用いられる。これは、学習した予測モデルが上手く予測できなかった学習データに重みをつけて更に学習を繰り返すことによって作成した複数の予測モデルを組み合わせる方法である。LightGBMにはLeaf-Wiseが採用されており、大規模データの学習に強く、モデルの学習速度が速いことが特徴である。SHAP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。SHAP値が大きいほど予測における寄与度が高いことを示す。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。

解析の手順は次のとおりである。まず先行研究が提案した3480ケースの断層シナリオのすべてについて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。予測対象地域は徳島県阿南市周辺とした。西南日本沖の海底にすでに設置されている海底水圧計の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と標準偏差を目的変数として、LightGBMにより津波浸水予測モデルを構築した。予測モデルの検証のため、別の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。

予測値と真値のRMSEは約0.6mであり、浸水分布をほぼ再現できた。各観測点のSHAP値から、予測地点から比較的近い観測点が津波浸水予測に最も寄与度が高いことが分かった(図1)。一方で、予測地点から離れた地域でも、SHAP値が周辺より高い観測点がまばらに存在している。四国西部～九州沖では観測点2点が周囲の観測点と比較してSHAP値が高かった。

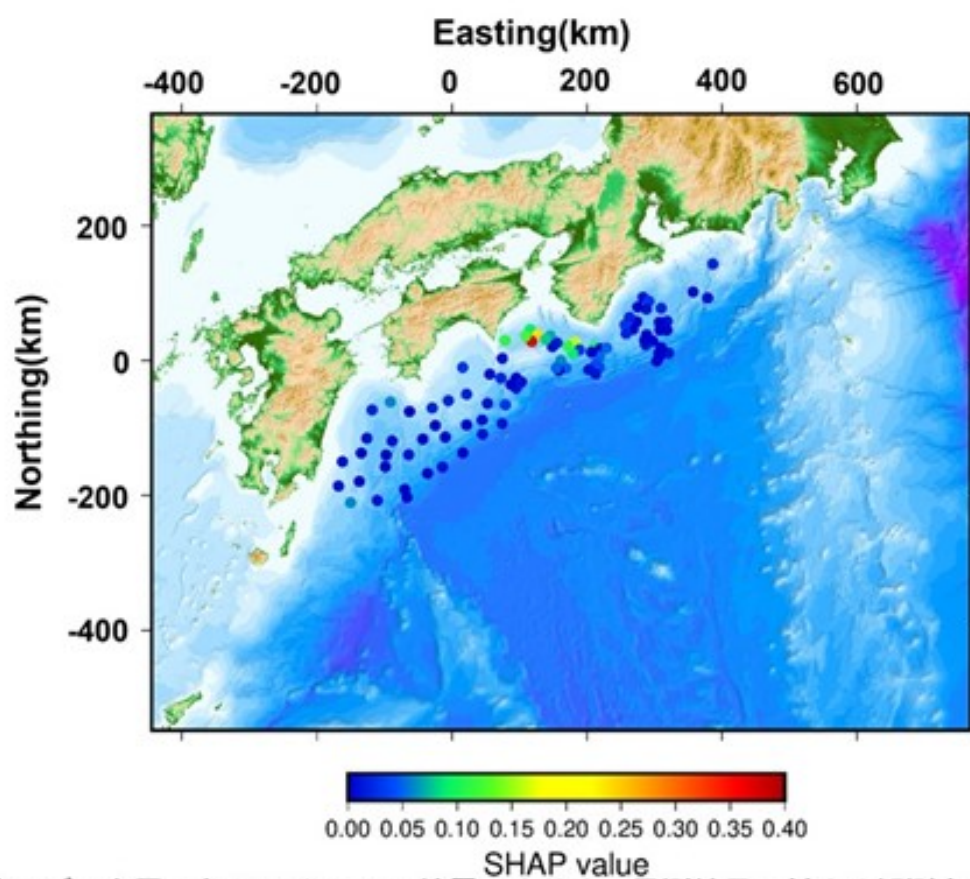


図1. 構築モデルを用いた3480ケースの地震シナリオの予測結果に対する観測点91地点の寄与度