

機械学習による氾濫推定モデルの中小河川への適用性の検討

九州産業大学建築都市工学部 学生会員 石田竜也 九州産業大学建築都市工学部 非会員 吉田昂洋
九州産業大学建築都市工学部 正会員 佐藤辰郎

1. はじめに

近年、自然災害をもたらす異常気象が毎年のように発生し、全国各地で人的被害が報告されている。災害による人的被害を軽減するためには、河川水位のみならず、氾濫解析により堤内地の浸水範囲や浸水深を予測することが重要である。従来の基礎方程式をベースとした氾濫解析では、流域内の多数の箇所想定される越水・溢水地点および破堤点の計算に多く計算時間やコストを必要とする。そのため本研究では、一旦モデルが構築されれば迅速に予測結果を出力可能な機械学習に注目する。

これまで、一級河川等の大河川からの氾濫予測に対しては、機械学習モデルの適用性が一定程度示されてきたものの、多くの河川への汎用性、特に人的被害が多発し、氾濫予測の必要性の高い中小河川への適用性は明らかとなっていない。本研究では、複数河川の教師データを基に機械学習による氾濫推定モデルを作成した上で、大河川と中小河川におけるモデルの精度を検証する。

2. 機械学習による氾濫推定モデルの構築

菅原らの研究¹⁾を参考に、機械学習を用いて、浸水深を目的変数、地形指標や氾濫流量を説明変数とする回帰モデルを作成した。

2.1 教師データ

物理モデルによる氾濫解析を実施し、教師データの作成を行った。氾濫解析の対象河川は、荒川・加古川・鬼怒川・高川・利根川・由良川である。荒川:60 地点、加古川:56 地点、鬼怒川:53 地点、高川:16 地点、利根川:65 地点、由良川:39 地点の合計 289 地点の破堤点を設定した。289 地点の破堤点それぞれに対して、8 つの流量ケースで平面 2 次元氾濫計算を行い、メッシュ毎の浸水深データを目的変数の教師データとして整備した。氾濫計算は、DioVISTA/Flood (日立パワーソリューションズ) を用い、パラメータは浸水想定区域図作成マニュアル(第 4 版)に準拠した。また、説明変数には浸

水候補地の「傾斜角を重みとしたコスト距離」、「傾斜角」、「曲率」、「標高」、「破堤点を基準とした相対標高」、「TRI」、「TWI」、「SPI」、「破堤点からの距離」、「土地の利用状況」、「河川からの距離」の 11 変数及び破堤点の「累積氾濫流量」、「破堤時間」を加えた計 13 変数を指定した²⁾。

2.2 機械学習モデル

勾配ブースティングフレームワークである LightGBM をモデルとして採用した。Gradient Boosting Decision Tree (GBDT) は多くのエンジニアリングの最適化に採用されており、その中でも LightGBM モデルは、次元・データサイズが大きいという、これまでの XGBoost 等の GBDT モデルの欠点を改善し、殆ど同精度でありながら学習過程を 20 倍以上効率化している²⁾。本研究において使用する教師データは表データであることから、LightGBM とは好相性であると考えられる。

2.4 モデルの学習

損失関数は、回帰分析に頻繁に用いられる Root Mean Squared Error (RMSE) を使用した。本研究では、学習に使用するデータを 2000 万個用意し、訓練データと検証データを 7:1 の割合に分割し学習を行った。lightGBM の主要なハイパーパラメータである決定木の個数を設定する num_iterations を 2,000、決定木の葉数の最大数を設定する num_leaves を 5,000 とした。また、学習に使用した浸水が発生するデータと発生しないデータの割合は、1:2 である。

2.5 検証データによるモデルの評価

前述の条件でモデルを作成し、学習が終了した時点での検証データに対する RMSE は 0.155、決定係数 R^2 は 0.974 であった。学習過程で RMSE が収束に向かっていたこと、RMSE が 0.155 と低く、決定係数の値が 0.974 と 0.9 を超えるほどの値であることから検証データに対して高精度の予測ができています。

2.6 テストデータによるモデルの評価

学習過程に使用されていない未知のテストデータを各破堤点毎に 1 万個用意し、計 289 万のテストデータに

よる精度検証を行った。その結果を表 1 に示す。表 1 は「浸水が発生する場合を正」として計算している。テストデータの浸水が発生するデータと発生しないデータの割合は学習用データと同じ 1:2 とした。表 1 の RMSE や決定係数 R^2 から未知のデータに対しても RMSE は 0.19 程度の誤差しかなく、決定係数も 0.9 を超えていることから高精度の予測が行われているといえる。加えて、Recall, Specificity が 0.9 を超えていることから、真陽性や真陰性の予測精度が 9 割を超えるほど浸水予測は高精度である。

表 1 テストデータに対するモデルの精度

RMSE (m)	0.19	Precision	0.88
MAE (m)	0.08	Recall	1.00
R^2	0.95	Specificity	0.94

3. 中小河川への適用性の検証

栃木県永野川及びその支流を含む 7 破堤点(図-1)において DioVISTA/Flood により氾濫計算を行い、各破堤点毎に 1 万個の計 7 万個のテストデータを作成した。流量ケースは各破堤点毎の一つずつ設定している。テストデータの浸水が発生するデータと発生しないデータの割合は 1:2 とした。2 章にて作成したモデルを使用し、本テストデータに対して精度検証を行った結果を表 2 に示す。表 2 の結果から RMSE は 0.7791 と誤差が大きく、 R^2 の値からも浸水深の予測が正常に行われていない。また、Recall の値が低いことから真陽性の予測の精度も低い(図-2)。これらから、大河川のデータで構築したモデルを用いて、中小河川の浸水氾濫予測を実施することは難しいものと考えられる。

表 2 小中規模河川データに対するモデルの精度

RMSE (m)	0.78	Precision	0.81
MAE (m)	0.36	Recall	0.62
R^2	-0.34	Specificity	0.93

4. 結論

比較的規模の大きな 6 つの河川を用いて学習したモデルは、同河川の未知のデータに対しても優れた精度で予測が行えていることが確認できた。

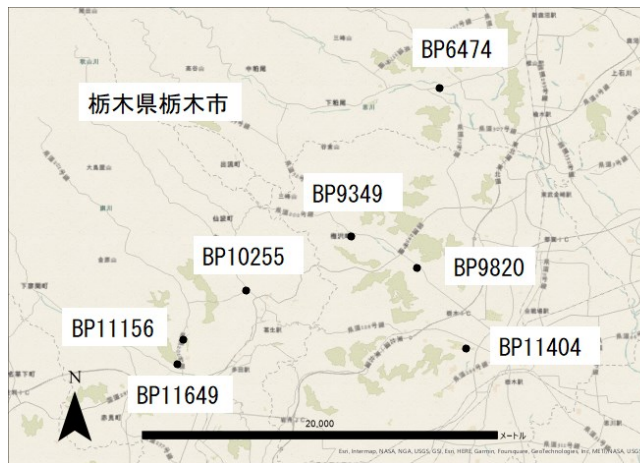
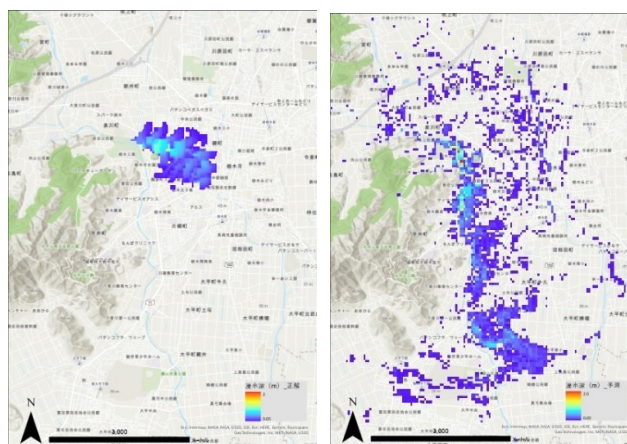


図-1 栃木県永野川と支流に設けた 7 破堤点



正解データ

予測結果

図-2 破堤点 BP11404 におけるモデルの予測結果

しかし、同モデルを中小河川に展開した時、十分な予測精度を得られないことが判明した。大規模河川と比較して氾濫流量等のスケールが大きく異なることが原因と考えられる。中小規模河川の浸水氾濫予測には、中小規模河川の学習データを用意する必要があるものと考えられる。

参考文献

- 1) 菅原ほか: 機械学習を用いた氾濫域推定モデルの地域間比較に関する研究, 河川技術論文集, 第 28 巻, pp.43-48, 2022
- 2) Guolin Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157, 2017