地形・地質・雨量情報を利用した機械学習による CL の土壌雨量指数の推定

九州大学大学院　学生会員　○竹田　梨夏
九州大学大学院　正会員　笠間　清伸
九州大学大学院　正会員　古川　全太郎

１．目的

近年、気候変動によって豪雨が増加傾向にあり、それに伴い土砂災害の激甚化が懸念されている。現在運用されている土砂災害警戒情報は、過去の自然災害をもとに発令基準が定められており[1]、降水量以外に土砂災害発生の原因となる地形や地質、植生等に関する情報が考慮されていないという問題点がある。そこで、降水量だけではなく土地の総合的な情報を考慮し、かつこれからの気候変動にも対応可能な、新しい基準の設定が必要とされる。本文では、福岡県内を対象に CL の横軸に相当する土壌雨量指数を 3 種の機械学習アルゴリズム (ランダムフォレスト (以下 RF) 、XGBoost および LightGBM) によって推定し、これらの精度を比較した。

２．解析概要

説明変数は**表 1** の通りである。各データは福岡県内の土砂災害が確認された 177 地点を対象として、1 km メッシュごとに収集した[2]。雨量情報以外のデータについては、メッシュごとに平均値、最大値、最小値、中央値、最頻値および標準偏差値等を算出・格納した。目的変数は CL 線の横軸に相当する土壌雨量指数である。説明変数を訓練データ (80 %) と検証データ (20 %) に分け、交差検証とグリッドサーチを用いて最適なパラメータの探索を行った。その後、訓練データを用いて学習を行った。学習後のモデルと検証データを用いて土壌雨量指数の予測を行い、訓練データによる予測結果と比較をした。また再帰的特徴量削減 (以下 RFE) を行うことで重要度の高い説明変数のみを用いて学習を行い、精度向上を試みた。

表 1　説明変数一覧

地形情報	標高、傾斜角、傾斜方向、流向、集水面積、TPI (地形起伏指数)、TRI (地形凹凸指数)
地質情報	地質の種別番号
雨量情報	各月の降水量 (平年値)、最大雨量 (1･2･3･6･12･24 時間)
その他	土地利用、植生群

計 177 地点・85 種類

３．解析結果

それぞれのモデルにおける決定係数は**表 2** の通りである。訓練データを用いて予測した結果では XGBoost で、検証データを用いて予測した結果では LightGBM でそれぞれ決定係数が最大になった。**表 3** は各モデルで説明変数の重要度を算出した結果、重要度が高かったものをまとめたものである。どのモデルでも、8 月の雨量平年値と最大 24 時間雨量が共通して重要とされていることが分かった。また雨量情報の重要度が高くなる傾向が見られた。**表 4** は、検証データを用いた予測結果の誤差が 10mm または 20mm 以内だった地点の割合をまとめたものである。また、**図 1～3** は予測値と実測値の関係をモデルごとにまとめた散布図である。散布図中の実線は実測値と観測値が等しい直線であり、点線はそれぞれ予測誤差が±10 mm および±20 mm

表 2　決定係数一覧

	RF	XGBoost	LightGBM
訓練データ	0.9282	0.9999	0.9951
検証データ	0.8517	0.8191	0.8615

表 3　各モデルでの説明変数の重要度

順位	RF	XGBoost	LightGBM
1	8 月の雨量平年値	8 月の雨量平年値	8 月の雨量平年値
2	最大 24 時間雨量	最大 24 時間雨量	最大 24 時間雨量
3	最大 1 時間雨量	7 月の雨量平年値	6 月の雨量平年値
4	1 月の雨量平年値	植生群最大値	7 月の雨量平年値
5	6 月の雨量平年値	植生群指標数	4 月の雨量平年値

表 4　誤差 10 mm または 20 mm 以内だった地点の割合

	RF	XGBoost	LightGBM
±10 mm 以内 (%)	86.11	77.78	77.78
±20 mm 以内 (%)	94.44	94.44	97.22

キーワード　機械学習，CL，地形情報，地質情報，雨量情報
連絡先　〒819-0385　福岡市西区元岡 744　W2-1110　TEL 092-802-3384

© Japan Society of Civil Engineers

の直線である。決定係数が小さくなった RF と XGBoost でも、予測結果の誤差が 20 mm を超えた地点は 10 %未満となった。

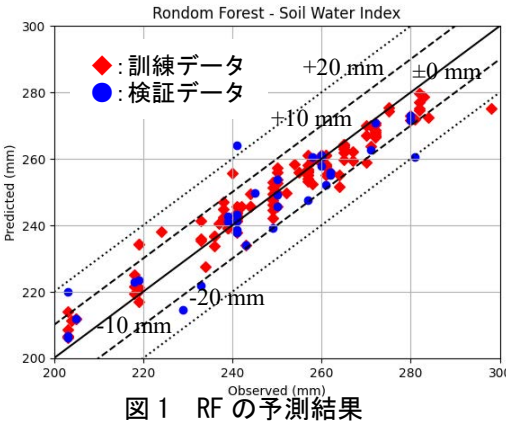

図 1　RF の予測結果

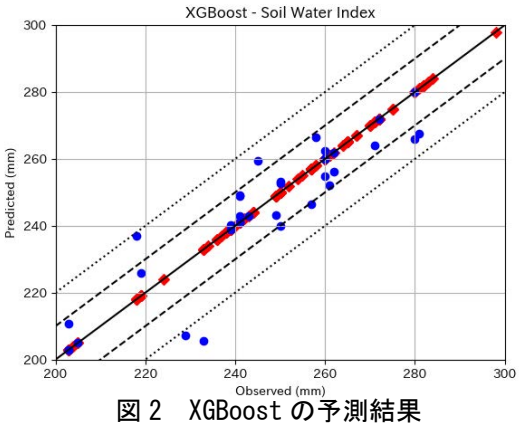

図 2　XGBoost の予測結果

4．RFE 結果

RFE を用いた学習の概要は以下の通りである。1) すべての説明変数を用いて学習を行う。2) 重要度の最も低い説明変数を削除する。3) 削除後の説明変数を用いて再び学習を行う。4) 説明変数の種類が目的の数 (今回は 40 種類) になるまで繰り返す。それぞれのモデルにおいて RFE を行った結果、抽出された説明変数は**表 4** の通りである。RF では雨量情報が、XGBoost と LightGBM では地形情報が比較的多く抽出された。5) 抽出された説明変数を用いて改めて学習を行う。上記の通り RFE を行った場合の、各モデルでの決定係数を**表 5** に示す。検証データを用いた推定では、LightGBM でのみ決定係数が上昇した。

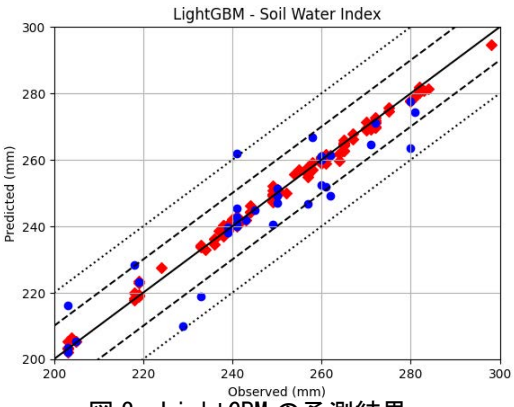

図 3　LightGBM の予測結果

表 5　RFE で抽出された決定係数一覧

	RF	XGBoost	LightGBM
地形情報	標高 (最小, 最頻値), 傾斜標準偏差, 傾斜方向 (合計, 平均, 最大, 最頻値, 標準偏差), TPI (合計, 中央, 最小値), TRI (平均, 最頻値), 集水面積指標数	標高 (合計, 中央, 最頻値, 標準偏差), 傾斜 (標準偏差, 最小値, 最頻値, 範囲), 傾斜方向 (合計, 平均, 最小, 最大, 最頻値, 標準偏差), TPI (中央, 最小値), TRI 最小値, 集水面積 (指標数, 標準偏差, 最大値)	標高 (標準偏差, 最小値), 傾斜 (標準偏差, 最大, 最頻値), 傾斜方向 (平均, 標準偏差, 最小, 最大, 最頻値), TPI (合計, 最小, 最大, 最頻値), TRI (標準偏差, 最大, 最頻値), 集水面積 (指標数, 合計, 最大値)
地質情報	種別番号平均値	種別番号平均値	種別番号平均値
雨量情報	雨量平年値(1~12 月, 1 年間), 最大雨量 (1, 2, 3, 6, 12, 24 時間)	雨量平年値 (1, 2~5, 7~9, 11 月, 1 年間), 最大雨量 (1, 2, 3, 24 時間)	雨量平年値 (3, 7, 8, 10~12 月) 最大雨量 (1, 2, 3, 6, 12, 24 時間)
その他	植生群 (指標数, 合計), 土地利用 (合計, 平均, 最小値, 標準偏差)	植生群 (指標数, 平均, 最大値, 標準偏差), 土地利用 (合計値, 標準偏差)	植生群 (指標数, 合計, 平均値, 標準偏差), 土地利用 (合計値, 標準偏差)

表 6　RFE 後の決定係数一覧

	RF	XGBoost	LightGBM
訓練データ	0.9308	0.9995	0.9942
検証データ	0.8364	0.7829	0.8792

5．まとめ

本文で得られた結果は以下の通りである。1) 今回用いた 3 つの手法の中では、検証データを用いた予測結果において最も決定係数が大きくなった LightGBM が適当であると言える。2) いずれのモデルにおいても、予測値の誤差が 20 mm 以上となった地点は全体の 10 %未満となった。3) RFE によって特徴量削減を行ったところ、学習手法によって抽出される特徴量が異なった。また、LightGBM では決定係数が上昇したが、RF と XGBoost とでは低下した。

謝辞：本研究は、JSPS 科研費 22K04318 の助成を受けたものです。ここに感謝の意を表します。

【参考文献】1) 気象庁, “土砂災害警戒情報・土砂キキクル（大雨警報（土砂災害）の危険度分布）”, 気象庁ホームページ, https://www.jma.go.jp/jma/kishou/know/bosai/doshakeikai.html (参照 : 2023-03-28). 2) 小野廉・石川達也・笠間清伸, 「降雨・地盤の地域特性を考慮した斜面災害リスク評価指標の提案」, 第 57 回地盤工学研究発表会講演集, DS-7-11, 2022.

© Japan Society of Civil Engineers