

機械学習を用いた土砂災害発生危険基準線の推定法に関する研究

九州大学大学院 学生会員 ○竹田梨夏 正会員 笠間清伸 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起これつつある。現在運用されている土砂災害警戒情報は、過去の斜面災害とその時の降雨情報をもとに発令基準が定められており、降雨以外に土砂災害発生に関係がある地形や地質、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、降雨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なる。本研究では、現行の CL 線と相関の強い情報を把握するため、福岡県内を対象に CL 線の縦軸に相当する 60 分間積算雨量を機械学習 (ランダムフォレスト (以下、RF)、XGBoost および LightGBM) によって推定し、精度等を比較・考察した¹⁾。

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。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等を算出・

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・93 種類

格納した。目的変数は CL 線の縦軸に相当する 60 分間積算雨量である。説明変数を訓練データと検証データに分け、訓練データを用いて学習を行った。学習後のモデルに対して、検証データを用いて 60 分間積算雨量の予測を行い、検証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 (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²⁾。

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 LightGBM で、検証データを用いて予測した結果では XGBoost で、それぞれ決定係数が最大になった。図 1~3 は予測値と実測値の関係を各モデルでまとめた散布図である。横軸に実測値を、縦軸に予測値をプロットしている。図中の直線について、中央は予測値と実測値が等しい場合、上下は予測誤差が ±5 mm の場合を表している。赤いマークは訓練データ、青いマークは検証データの結果を示している。検証データによる予測結果の決定係数が他のモデルより小さくなった RF でも、検証データによる予測結果の内 90.74 %が ±5 mm に収まった。表 3 はそれぞれのモデルにおける説明変数の重要度上位 5 種類を示す。すべてのモデルにおいて、6 月の雨量平年値と地質番号の平均値が重要と

表 2 各モデルでの決定係数

	RF	XGBoost	LightGBM
訓練データ	0.8978	0.9684	0.9957
検証データ	0.5011	0.7067	0.6505

表 3 それぞれのモデルでの説明変数の重要度

順位	RF	XGBoost	LightGBM
1	最大 24 時間雨量	地質種別番号の平均値	6 月の雨量 (平年値)
2	8 月の雨量 (平年値)	9 月の雨量 (平年値)	最大 3 時間雨量
3	6 月の雨量 (平年値)	最大 6 時間雨量	2 月の雨量 (平年値)
4	2 月の雨量 (平年値)	6 月の雨量 (平年値)	地質種別番号の平均値
5	地質種別番号の平均値	土地利用種別番号の合計値	8 月の雨量 (平年値)

されていることが分かった。その他の重要度の高い説明変数については雨量に関するものが多く見られた。

4. RFE 結果

RFE を用いた学習の流れ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った。2) 重要度の最も低い説明変数を削除した。3) 削除後の説明変数を用いて再学習を行った。4) これを説明変数の種類が 20 種類に達するまで繰り返した。それぞれのモデルにおいて RFE を行い、抽出された説明変数は表 4 の通りである。5) 20 種類の説明変数を用いて、改めて学習を行った。RFE を行った場合の各モデルでの決定係数を表 5 に示す。検証データによる予測の場合、RF と LightGBM では決定係数が大きくなったが、XGBoost では小さくなった。

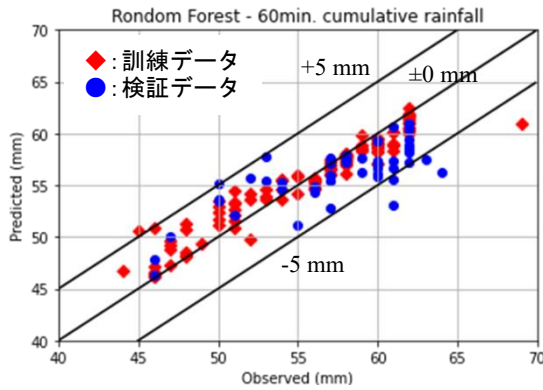


図 1 RF の予測結果

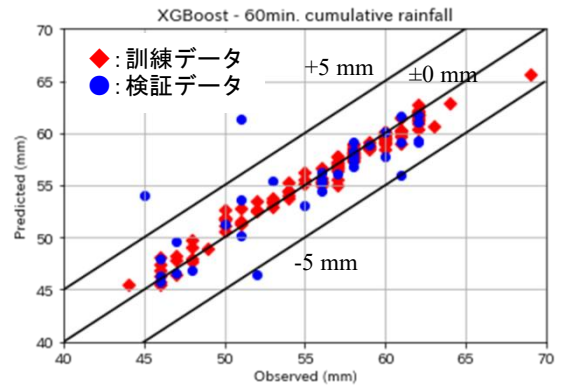


図 2 XGBoost の予測結果

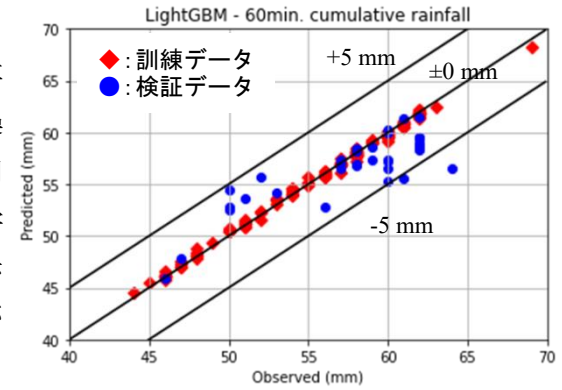


図 3 LightGBM の予測結果

表 4 それぞれのモデルでの RFE 後の説明変数一覧

	RF	XGBoost	LightGBM
地形情報	傾斜方向中央値 集水面積 最大雨量 (1, 2, 3, 12, 24 時間)	傾斜 (平均, 最小値) 標高最頻値 傾斜方向 (平均, 中央値) 集水面積、粗度	標高最頻値 傾斜方向 (合計, 平均, 中央, 最大値, 標準偏差) TPI (中央, 最頻値, 変動係数) TRI 標準偏差 集水面積最大値
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 (3, 12 月以外) 最大雨量 (1, 2, 3, 12, 24 時間)	雨量平年値 (4-9 月) 最大雨量 (1, 3, 12, 24 時間)	雨量平年値 (6, 8 月) 最大 3 時間雨量
その他	植生群 (平均値, 範囲)	植生群 (合計, 標準偏差)	無し

表 5 RFE 有の場合の各モデルでの決定係数

	RF	XGBoost	LightGBM
訓練データ	0.8979	0.9810	0.9910
検証データ	0.5935	0.6609	0.6702

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決定係数が最も大きくなった XGBoost が適当と言える。2) 決定係数が最も小さくなった RF についても、予測値が実測値より 5 mm 以上変動した地点は全体の 10%未満だったため、実際の降雨の様子を考慮すると一定の実用性があると言える。3) RFE によって特徴量削減を行ったところ、RF と LightGBM では決定係数が上昇したが、XGBoost では決定係数が低下した。今後、土壌雨量指数についても推定モデルの作成を行い、同様にモデルによる精度の違いや説明変数選択について検討していく予定である。

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) 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予稿集, 2022
- 2) scikit-learn developers (BSD License), “sklearn.feature_selection.RFE”, scikit learn, https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html (参照 : 2023/1/5)