

B-03

機械学習に基づく豪雨による土砂崩壊発生地点の予測

17T0221F 桑折 奎吾

指導教員：丸山 喜久 劉 ウェン

1. 研究背景と目的

平成30年7月豪雨（西日本豪雨）は、全国で死者・行方不明者270名以上、全壊の住家6,700棟以上の甚大な被害をもたらした¹⁾。広島県では、死者・行方不明者114名のうち、約8割の87名が土砂災害の犠牲者とされている（2018年9月時点）²⁾。

気象庁によると、平成30年7月豪雨に代表されるような激甚な被害を与える豪雨の発生は、長期的な視点では増加傾向にある³⁾。そのため、特に大きな被害をもたらす土砂災害の被害を減少させるために、土砂崩壊発生地点を予測することは重要である。

そこで本研究では、平成30年7月豪雨による土砂災害の被害を受けた地域を分析し、土砂崩壊の発生と関係のある地形・地質情報と降雨情報を説明変数とし、機械学習によって土砂崩壊発生予測モデルの構築を目標とする。

2. 使用データと対象地域

使用するデータは、国土地理院が公表している崩壊地等分布図⁴⁾と、国土数値情報⁵⁾、国土調査⁶⁾、微地形区分（5次メッシュ）⁷⁾、土壌雨量指数⁸⁾である。崩壊等分布図は、平成30年7月豪雨により土砂崩壊が発生したとみられる地点を航空写真から判別したものである。本研究では、そのうち広島、広島東部のデータを使用した。国土数値情報では、標高、傾斜角などの地形データが5次メッシュ（250mメッシュ）単位で集約されている。国土調査のデータは、土壌、表層地質など地質のデータが格納されている。土壌雨量指数⁹⁾は、降雨が土壌中に水分量としてどれだけたまっているかを、タンクモデルを通じて定量化したもので、5kmメッシュごとに30分刻みの土壌雨量指数が記録されている。今回は広島県に大雨特別警報が発令されていた平成30年7月6日～7日のものを使用する。土壌雨量指数の最大値、30分毎の差分の最大値を説明変数として採用した。

使用データはGISを用いて、全てのデータを5次メッシュ単位に集約してから分析を行った。土壌、表層地質のデータについて、1つの5次メッシュに複数種含まれる場合は、面積の割合が大きいものをそのメッシュの土壌、表層地質データとした。

3. 機械学習に用いる説明変数の選択

機械学習の説明変数を選択するために、各説明変数と土砂崩壊発生率との相関関係を確認する。土砂崩壊発生率は、土砂崩壊地点数をメッシュ数で除した値として定義した。図-1に、土壌雨量指数の最大値と土砂崩壊発生率の関係を示す。



図-1 土壌雨量指数の最大値と土砂崩壊発生率の関係

全ての説明変数について土砂崩壊発生率との相関関係を確認し、相関が全く見られないものは説明変数として採用しないこととした。この結果、機械学習の説明変数として採用したものは、最高標高、最大傾斜角、土壌雨量指数の最大値、土壌雨量指数の30分毎の差分の最大値、微地形区分、土壌、表層地質である。

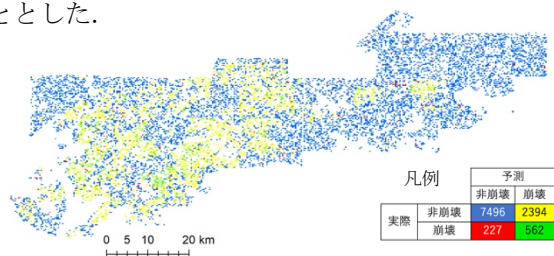
4. 機械学習による予測モデルの構築

前章で採用した説明変数を用いた機械学習により土砂崩壊発生予測モデルを構築し、最も精度の高いモデルを模索した。モデルの説明変数は3つ以上とし、微地形と土壌、表層地質については、カテゴリ変数であるためダミー変数として扱う。まず、カテゴリ変数を用いずに予測モデルを構築し、最も精度の高いモデルに対し、カテゴリ変数を導入することとした。

本研究では、崩壊地点を含む5次メッシュを「崩壊メッシュ」、含まない5次メッシュを「非崩壊メッシュ」とした2クラス分類をランダムフォレストによって行った。ランダムフォレストは、複数の決定木を弱学習器として用いるアンサンブル学習の一つである¹⁰⁾。個々の決定木は高い識別性能を持たないが、それらを複数用いてそれぞれの結果を補うことによって高い予測性能を得ることができる。この手法の特徴としては、説明変数の結果に対する影響度が算出できること、効率が良く計算速度が高速であることなどが挙げられる。モデルの性能を大きく左

右するハイパーパラメータをグリッドサーチ（交差検証を含む）に基づき設定した。本研究では、崩壊メッシュの F 値を評価指標とし、6 種類のハイパーパラメータを設定した。また、全データ数に対し、70%を学習データ、30%をテストデータとした。

図-2 に、テストデータに対して、崩壊メッシュの F 値および再現率が最も良好なモデルによる分類結果を示す。このモデルの F 値、適合率はそれぞれ 0.30 と 0.19 である。また、再現率 (0.71)、正答率 (0.75) もあまり良い結果ではない。これには、図-2 の右上のように崩壊メッシュが全く予測されない地域があることが影響していると考えられる。この問題を解消するために、崩壊メッシュのデータを精査することとした。

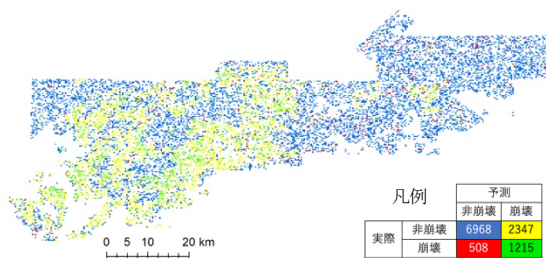


説明変数：最高標高，最大傾斜角，土壌雨量指数の最大値，土壌雨量指数の 30 分毎の差分の最大値

F 値：0.30，再現率：0.71，適合率：0.19，正答率：0.75

図-2 機械学習による土砂崩壊発生地点の予測結果

そこで、同じ地域で広島大学が収集した平成 30 年 7 月豪雨によって発生した土砂崩壊発生地点のデータ¹¹⁾を新たなデータとして用いることとした。国土地理院と広島大学のデータのどちらかで土砂崩壊とみなされるメッシュを崩壊メッシュとした。これにより崩壊メッシュ数は約 2.14 倍となった。図-3 にデータ変更後の最も精度の高いモデルの結果を示す。



説明変数：最高標高，最大傾斜角，土壌雨量指数の最大値，土壌雨量指数の 30 分毎の差分の最大値，微地形区分

F 値：0.46，再現率：0.71，適合率：0.34，正答率：0.74

図-3 データ変更後の土砂崩壊発生地点の予測結果

データ変更前と比較すると、F 値と適合率は上昇した。しかし、データ変更後も再現率は 0.71 で変化しなかった。図-2 と図-3 によると、モデルが崩壊メッシュを見逃す地域が比較的似ているため、精度の向上を図るには、説明変数の見直しや土砂崩壊の形態の考慮などが必要と考えられる。

5. まとめと今後の課題

本研究では、ランダムフォレストを用いて土砂崩壊発生地点の予測モデルの構築を行った。国土地理院のみの崩壊地点のデータを使用した予測モデルは、良好な分類結果を示さなかった。そこで、広島大学による崩壊地点データを加えモデルを再構築したところ、F 値と適合率は上昇したが、再現率は 0.71 で変化しなかった。今後は、説明変数を増やすほか、2 段階の機械学習を行うなど、データと手法の両側面から改善方法を探り、精度の良いモデルを構築することを目標とする。また、他の地域に適用し、汎用性も確認していく予定である。

参考文献

- 1) 消防庁：平成 30 年 7 月豪雨及び台風第 12 号による被害状況及び消防機関等の対応状況（第 60 報），2019
- 2) 海堀正博，長谷川祐治，山下祐一，崎田博史，中井真司，桑田志保，平松晋也，地頭園隆，井良沢道也，清水収，今泉文寿，中谷加奈，柏原佳明，加藤誠章，鳥田英司，平川泰之，吉永子規，田中健路，林拙郎：平成 30 年 7 月豪雨により広島県で発生した土砂災害，砂防学会誌，Vol.71，No.4，p.49-60，2018
- 3) 気象庁：気象業務はいま 2020 | 特集 激甚化する豪雨災害から命と暮らしを守るために，
<https://www.jma.go.jp/jma/kishou/books/hakusho/2020/index1.html#toc-001>
- 4) 国土地理院：平成 30 年 7 月豪雨に関する情報，
<https://www.gsi.go.jp/BOUSAI/H30.taihuu7gou.html>
- 5) 国土交通省：国土数値情報ダウンロードサイト，
<http://nlftp.mlit.go.jp/ksj/>
- 6) 国土交通省：国土調査（土地分類調査・水調査），
<https://nlftp.mlit.go.jp/kokjo/inspect/inspect.html>
- 7) 若松加寿江，松岡昌志，久保純子，長谷川浩一，杉浦正美：日本全国地形・地盤分類メッシュマップの構築，土木学会論文集，No. 759/1-67，213-232，2004.
- 8) 株式会社ハレックス：DreamAll「土壌雨量指数」
- 9) 気象庁：土壌雨量指数，
<https://www.jma.go.jp/jma/kishou/known/bosai/dojoshisu.html>
- 10) 波部 斉：ランダムフォレスト，コンピュータビジョンとイメージメディア，Vol.2012-CVIM-182，No.31，p.1-8，2012
- 11) 広島大学平成 30 年 7 月豪雨災害調査団（地理学グループ）：平成 30 年 7 月豪雨による広島県の斜面崩壊分布図（第四報），2018-08-02