

土砂災害発生有無の分類における機械学習モデルの構築と発生要因の抽出

信州大学大学院 学生会員 ○佐藤拓馬
信州大学工学部 正会員 小山 茂

1. はじめに

日本では、土石流やがけ崩れ等の土砂災害が発生することにより、これまでに数多くの被害を経験してきた。これは、急傾斜が多い山岳地形や脆弱な地質といった地形的特徴を指す素因と、台風による大雨やゲリラ豪雨といった災害の引き金となる誘因が日本においては避けられないためである。

土砂災害は住民の生命や財産を脅かすものであり、被害を最小限に抑えるには早期の避難が重要となる。そのためには、任意地点における災害発生の危険性を把握することが必要になるが、土砂災害は素因と誘因の複雑なメカニズムにより発生するため、その判断を人間が下すのは困難な場合がある。

本研究では、任意地点の土砂災害の危険性を評価するため素因と誘因に着目し、人間の理解が困難な規則やパターンを見出すことに優れた機械学習を、土砂災害発生有無の判断に適用する。過去に土砂災害が発生した長野県下の事例に対して、機械学習の一手法である勾配ブースティング木を用いて学習器を構築し、ある地点の地形情報や降雨がどの程度災害発生の引き金と成り得るかの判断を試みる。

2. 研究手法

(1) 研究の対象

本研究で対象とする土砂災害は、表層崩壊である土石流、がけ崩れとする。また、機械学習を行うにあたっては、土砂災害が発生した地点での災害、地形、降雨情報を用いる。

(2) 機械学習

機械学習とは、データの構造に関する理論や関係性が明白でなくとも、コンピュータを用いデータを精査することで、その構造を探ることが出来るという手法である。各データを構成する説明変数に対応するデータを特徴量、その説明変数に対応する目的変数をラベルと呼ぶ。データを構成する様々な特徴量とラベルの関係から、パターンや法則を見つけて学

習器を構築する。この学習器にある特徴量を含むデータを入力すると自動で判断してラベルを出力する。機械学習で扱うのは、ラベルが所属する集合を予測する分類と、ラベルの値を予測する回帰があるが、本研究は各特徴量を基に土砂災害の発生有無を推測するので、分類問題に対応する。

(3) 研究に用いる情報

長野県における2004年から2011年までの8年間の土砂災害データから得られた108の発生データに加えて、任意に選んだ324の非発生データの合計432データを用いる。発生データ内の災害種の内訳は、土石流が32、がけ崩れが76データとなっている。

国土地理院の場所情報コード API¹⁾を用い、任意地点とその周囲8点の情報を得る。各地点について、標高差平均、標高差の絶対値の平均、傾斜角平均、最大傾斜角、断面曲率、平面曲率、曲率、最大曲率を算出する。更に、気象庁の気象業務支援センターが刊行している1kmメッシュ解析雨量データ²⁾から任意地点の座標に該当するメッシュの解析雨量を用い、災害発生推定時刻の前6時間累積雨量、12時間累積雨量、24時間累積雨量、48時間累積雨量、72時間累積雨量を算出し、以上8つの素因と5つの誘因を特徴量とする。一方、学習器の出力となるラベルとして災害発生有無を与える。

3. 機械学習における学習器の概要

(1) 決定木

決定木とは、木構造を用いて分類を行う機械学習手法の1つである。学習用データ1つの特徴量に対して条件を与えることで、その条件の真偽に対応して2つのノードに分岐させる。この操作を、ノード内データが単一のラベルとして分類できるまで繰り返すことで決定木が構築される。

(2) 勾配ブースティング木

一般に、木構造1つから成る決定木は、汎化性能が優れていないという問題がある。性能の高い学習

器を得るために複数の決定木を構築して、それらの推測結果を統合し、1つの推測結果を出力する手法として勾配ブースティング木³⁾がある。勾配ブースティング木は多数の決定木を構築して結果を予測するアンサンブル学習の一種であると同時に、学習の際に、順に決定木を構築する中で一つ前の決定木の誤差を次の決定木が修正していくブースティングという手法も行っている。決定木よりも勾配ブースティング木を用いるほうが高い精度を期待できるため、本研究ではこちらを採用する。

(3) 研究の流れ

まず始めに、土石流では全356データ、がけ崩れでは全400データからそれぞれ95.5%を学習用データとして無作為に取り出し、表-1のように分割する。この学習用データに対してそれぞれ勾配ブースティング木を構築する。残る4.5%のデータを検証用データとして構築した学習器に入力し、そのデータが土砂災害発生か非発生かを出力させる。以上のプロセスを1回とし、この操作を1000回繰り返すことで学習器の性能を確認する。

表-1 機械学習に用いるデータの内訳

	学習用	検証用	計		学習用	検証用	計
土石流	30	2	32	がけ崩れ	73	3	76
非発生	309	15	324	非発生	309	15	324
計	339	17	356	計	382	18	400

4. 結果

(1) 災害種別での出力結果の比較

災害種別に行った1000回の学習に対する各評価指標の算術平均値を図-1に示す。accuracy（正解率）は発生と非発生を正しく分類した割合、precision（精度）は発生と分類した結果のうち実際に発生であった割合、recall（検出率）は実際の発生を正しく分類できた割合を表す。F1値は精度と検出率の調和平均から算出され、学習器の性能を評価するために有効な指標である。図-1より、正解率はどちらも80.0%近くとなり未知のデータを比較的正しく分類可能である。ただ、どちらの災害種も精度と検出率が互いに低い水準にあり、災害の発生を正しく分類するのが困難なことが分かる。学習器の性能の改善を図るために、特徴量である素因の増加や累積雨量だけでなく実行雨量も加えることが必要だと考えている。

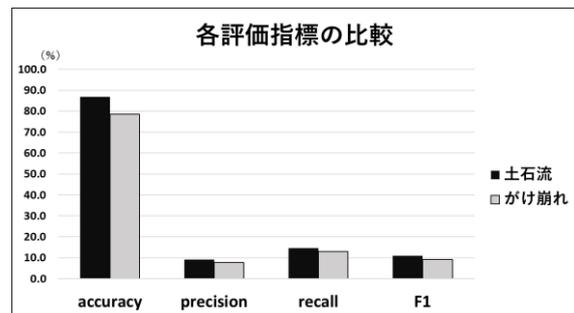


図-1 災害種別の各評価指標

(2) 各特徴量の重要度

勾配ブースティング木は、学習時に学習器を構築すると同時に、推測する際に各特徴量がどの程度影響を与えるかという重要度の算出が可能である。災害種別に試行した1000回の機械学習から得られた特徴量重要度の算術平均値を図-2に示す。この図から、土石流においては、誘因である24時間累積雨量、12時間累積雨量が災害発生に最も関与していることが明らかである。がけ崩れでは、傾斜角平均、断面曲率、平面曲率といった地形の形状に関する素因が土石流よりも重要視されている。誘因に着目すると、直前の累積雨量よりも長時間の累積雨量が推測に影響を与えていることが分かる。

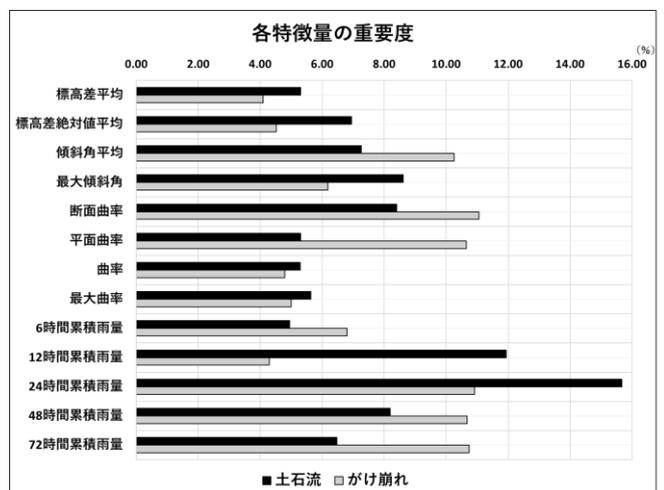


図-2 災害種別の各特徴量重要度

参考文献

- 1) 国土地理院, 場所情報コード API, 2020
https://ucopendb.gsi.go.jp/ucode/help_with_API.html
- 2) 国土気象庁, 気象業務支援センター, 1km メッシュ解析雨量 (DVD-ROM, 2006-2013)
- 3) 柴原一友, 築地毅, 古宮嘉那子, 宮武孝尚, 小谷義行: 機械学習教本, 森北出版, 2019.