

機械学習を用いた地形情報による土砂災害発生有無の分類

信州大学大学院 学生員 ○佐藤拓馬

信州大学工学部 正会員 小山 茂

信州大学工学部 正会員 大上俊之

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が災害発生の素因と成り得るのかどうかを判断することを試みる。

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、がけ崩れとし、地下水の影響が強い地すべりは除外する。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情報と地形情報を用いる。

(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て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70の発生データに加えて、任意に選んだ32の非発生データの合計102データを用いる。

国土地理院発行の数値地図標高 5m メッシュのDEMデータ¹⁾を用い、各地点について起伏、断面曲率、平面曲率、傾斜度を算出する。更に、国土交通省の20万分の1 土壤分類基本調査GISデータ²⁾と、土地細分メッシュデータ³⁾から表層地質、土壤分類、土地利用を調べ、以上7つの素因を特徴量とする。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。特徴量とラベルの例を表-1に示す。

表-1 データ内の特徴量とラベル

特徴量								ラベル
No.	表層地質	土壤分類	土地利用	傾斜	起伏	平面曲率	断面曲率	災害
1	未固結	黒ボク土	森林	18.7	228	0.181	0.275	○
2	固結	赤黄色土	森林	20.2	186	0.528	0.069	○
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
102	火山性	赤黄色土	田	0.23	70	1.08	0.667	×

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。図-1のように、学習用データの

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。

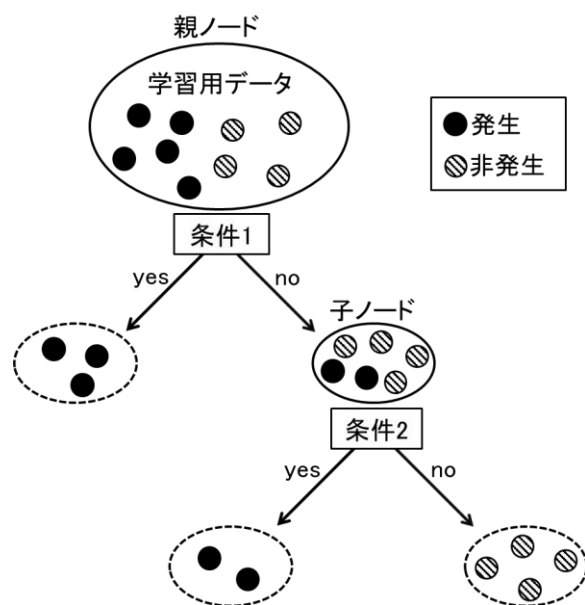


図-1 決定木構造のイメージ

(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⁴⁾がある。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、全102データから学習用データとして97データを無作為に取り出す。この学習用データに対して勾配ブースティング木を構築する。残る5データを検証用データとして構築した学習器に入力し、土砂災害発生となるか非発生となるのかを出力させる。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。

4. 結果

(1) 出力結果と実際の結果の比較

学習器の検証用データに対する出力と実際の災害発生有無が、いくつ一致したかを正解数とする。ま

た、学習器が非発生と判断したにも関わらず、実際には災害が発生した割合を偽陰性率とし、これらをまとめたものを表-2に示す。正解数の算術平均値は4.4と高く、未知データを比較的正しく分類できることが分かる。一方で偽陰性率は12.6%となり、場合によっては学習器が危険な判断をしてしまう結果となる。学習器の性能の改善を図るために、特徴量に誘因である降雨情報を追加することや、学習用データの増加が必要であると考えている。

表-2 正解数と偽陰性率の結果

正解数	5	4	3	2	1	0
回数	526	375	83	14	2	0
偽陰性率(%)	0.0	21.6	42.4	60.4	80.0	0.0

(2) 各特徴量の重要度

勾配ブースティング木は、学習器を構築すると同時に、推測する際に各特徴量が推測にどの程度影響を与たかという重要度の算出が可能である。正しく分類できた526回の学習器の結果から求めた値を表-3に示す。この表から、表層地質や土壌分類、土地利用といったある地点に与えられた情報より、起伏や曲率等の形状に関する地形情報が災害発生有無に重大な影響を与えていることが分かる。

表-3 各特徴量の重要度

	表層地質	土壌分類	土地利用	傾斜	起伏	平面曲率	断面曲率
重要度(%)	3.6	1.4	3.3	39.0	13.2	11.9	27.6

参考文献

- 1) 国土地理院基盤地図情報ダウンロードサービス：数値地図5mメッシュ（標高），2013.
<http://fgd.gsi.go.jp/download/>
- 2) 国土交通省国土政策局国土情報課の20万分の1土壌分類基本調査GISデータ，2011.
<http://tochi.mlit.go.jp/kihon/tochi-bunrui>
- 3) 国土交通省国土計画局，土地細分メッシュデータ，2006.
<http://nlftp.mlit.go.jp/ksj/gml/cgi-bin/download.php>
- 4) 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.