

地形情報に着目した土砂災害発生有無の分類における機械学習の適用

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

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にまずは素因に着目し、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) 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土壌分類基本調査GISデータ²⁾と、土地細分メッシュデータ³⁾から表層地質、土壌分類、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。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。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	土壌分類	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、機械学習手法の一つである。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。

(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、学習用データに強く依存するため、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

キーワード 土砂災害、地形情報、機械学習、勾配ブースティング、決定木

連絡先 〒380-8553 長野県長野市若里 4-17-1 信州大学工学部 TEL 026-269-5051

スティング木⁴⁾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。また学習の際に、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を行っている。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、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。評価する際の災害の発生、非発生の扱いはそれぞれ陽性、陰性とした。

4. 結果

(1) 出力結果と実際の結果の比較

1000 回の学習に対する各評価指標の算術平均値を表-2 に示す。正解率 (accuracy) は発生と非発生を正しく分類した割合、精度 (precision) は発生と分類した結果のうち実際に発生であった割合、検出率 (recall) は実際の発生を正しく分類できた割合を表す。F 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。表-2 より、正解率は 0.88 と高く、未知のデータを比較的正しく分類できる。また、精度、検出率が互いに高い水準にあり、F 値も 0.89 と、1 に非常に近い値なので汎化性能の優れた学習器を構築できたと考える。

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-3 に示す。偽陰性率の平均値は 0.13 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。構築す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と考えている。

表-2 出力結果に対する評価指標

正解率	精度	検出率	F 値
0.88	0.87	0.95	0.89

表-3 正解数と偽陰性率の結果

正解数	5	4	3	2	1	0
回数	539	338	105	17	1	0
偽陰性率	0.00	0.21	0.43	0.67	0.67	0.00

(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 539 回の学習器の結果から求めた値を図-1 に示す。この図から、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。

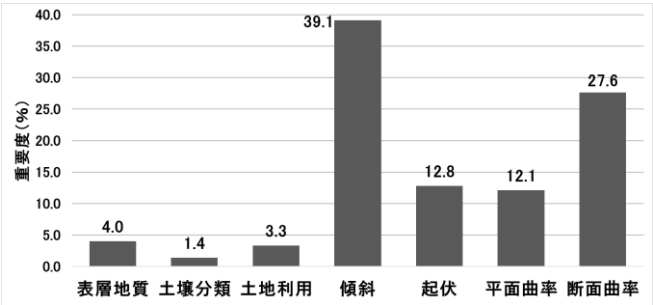


図-1 各特徴量の重要度

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.mlitt.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.mlitt.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.