

山斜面の傾斜角特徴を用いた教師なし外れ値検出による土砂災害危険箇所の識別

Discrimination of Landslide Hazard Areas by Unsupervised Outlier Detection Using a Feature of Mountain Slope Inclination

本田美矩[†]
Miku Honda

花沢明俊[†]
Akitoshi Hanazawa

1. はじめに

日本は豊かな自然に恵まれた国であるが、国土の7割が山地であるため河川は急勾配で流れも速く氾濫などが起きやすい地形である。また、活火山が多く分布しており地殻変動によって複雑で不安定な地形や地質が形成される。このような地形や地質などの条件に加えて、梅雨や台風などによる大雨や地震により自然災害が発生しやすい[1]。さらに近年は、ゲリラ豪雨や大型の台風に見舞われるなどの異常気象の増加により自然災害が多発している。その中でも土砂災害では、多数の死者・行方不明者がでている[2]。平成30年の豪雨による広島県での土砂災害や、令和2年の豪雨による熊本県での土砂災害はまだ記憶に新しい。そこで、土砂災害危険箇所の識別が必要不可欠であると考えられる。本研究では、山の斜面の傾斜角データを特徴量として機械学習を用いることで土砂災害危険箇所を識別する手法に着目した。

2. 目的

以前より機械学習を用いた土砂災害発生箇所の識別が行われている。従来の手法は、土砂災害発生箇所と非発生箇所のデータを使用し教師あり学習を用いることで土砂災害危険箇所を識別している。しかし、非発生箇所のデータには将来的に発生する可能性がある未発生箇所を利用しており信頼度が低いという課題があげられる。そこで本研究では、災害が発生した箇所のデータのみを使用して学習する教師なし外れ値検出を用いて行うこととした(図1)。

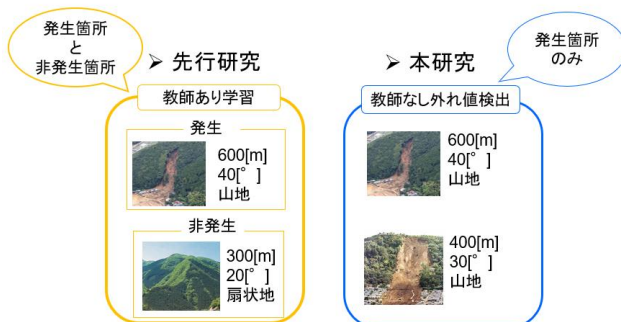


図1 専攻研究と本研究の比較

3. 関連研究

古川昭太・丸山喜久は、機械学習を行う前に共分散構造分析を行うことにより、地すべりに大きな影響を与える素因を「標高・傾斜角・傾斜方向・微地形」に絞り込み、特に標高と傾斜角の影響が大きいという結果を得た。その後、

教師あり学習である SVM と RF を適用して、地すべりや土砂崩壊発生地点を予測するモデルを構築した[3]。

石井唯嵩らは、教師あり学習である U-Net を用いて地形データである数値標高モデルと土石流氾濫開始地点データから土石流氾濫開始地点の推測を行っている[4]。

本研究では実際に傾斜角、土壌、地質、植生を識別データとして学習したところ傾斜角を含んだデータの精度が良いという結果が得られたため識別データとして傾斜角のみを使用した。

4. 実験方法

4.1 実験の流れ

実験の流れを図2に示す。まず、土砂災害発生箇所のデータと非発生箇所のデータを作成した。本研究では、平成30年の7月の豪雨で土砂災害が発生しなかった箇所を非発生箇所として用いた。次に、発生箇所のデータで教師なし外れ値検出を行い土砂災害発生箇所の識別モデルを作成した。最適なモデルを作成するために、交差確認法とグリッドサーチを用いた。本研究では、学習と最適化をモデル作成と定義している。最後に、モデル作成で使用しなかったデータを用いて識別実験を行った。

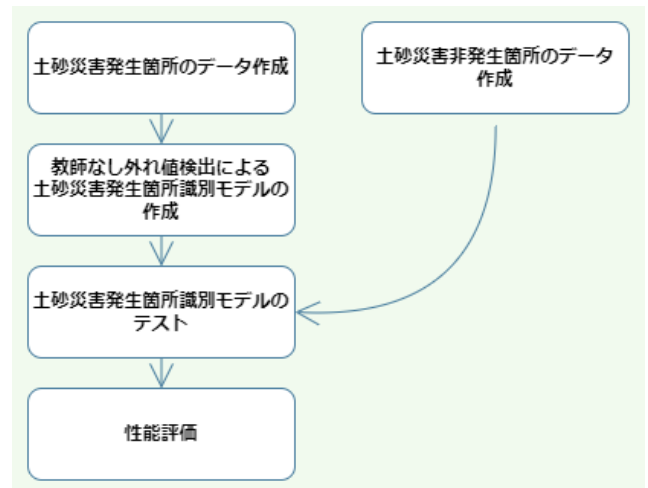


図2 実験の流れ

4.2 データ作成

土砂災害は、大きく「土石流」「地すべり」「がけ崩れ」の3つに分類することができる。本研究では、土石流とがけ崩れに限定して行うこととした。地すべりは地震などの影響で突然起きることもあるが、大抵は目に見えないほどゆっくりと移動していくためである[5]。対象地域は平成3

[†]九州工業大学 工学府
Graduate School of Engineering, Kyusyu Institute of
Technology

0年の豪雨により影響を受けた広島県東部とし、土砂災害が発生した箇所は国土地理院が提供している図3に示す崩壊地等分布図[6]を参照した。図3に含まれる赤い地点が崩壊地点である。

傾斜角データは、国土地理院が提供する数値標高データ(DEM)[7]より、プログラムを用いて算出した。DEMは、標高が5mメッシュごとに格納されているものを使用した。作成方法は、航空レーザー測量で標高精度は0.3m以内であり提供されているDEMデータの中で最も高い精度である。以下にデータの解析方法を示す。

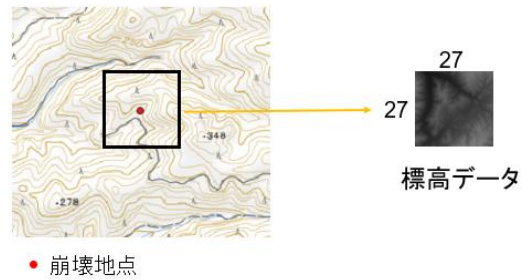


図4 データ構成

(1) 傾斜角データの算出

1. 地理情報システムの閲覧、編集、分析機能を有する QGIS を用いて国土地理院から取得した DEM データを、画像ファイル形式である TIFF に地理情報を付加したファイル形式である Geotiff 画像で出力する。
2. Geotiff 画像を、図4に示すように崩壊地点を中心に 27×27pixel に切り出す。1pixel に1つ標高が含まれるため、全部で 27×27 個の標高を含む。
3. 図5に示すように 3×3pixel の範囲で標高データの差分をとり x 方向と y 方向の傾斜を求める(図6上)。
4. 3 で求めた x ベクトルと y ベクトルの外積を計算することで法線ベクトル n を求める(図6中)。
5. xy 平面に n ベクトルを射影した n' ベクトルを求める。
6. n ベクトルと n' ベクトルがなす角 θ を求める。

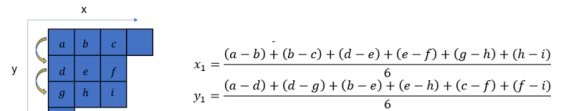


図5 傾斜の計算方法

$$\|n\| \|n'\| \cos \theta = \frac{1}{2} (\|n\|^2 + \|n'\|^2 - \|n - n'\|^2)$$

$$\theta = \cos^{-1} \left(\frac{\|n\|^2 + \|n'\|^2 - \|n - n'\|^2}{2 \|n\| \|n'\|^2} \right)$$

7. 90度からθを引くことで、傾斜角θ'を算出する(図6下)。

$$\theta' = 90 - \theta$$

(2) 特徴量の作成

(1)で算出した 27×27pixel の傾斜角を含む画像を9等分しそれぞれ傾斜角のヒストグラムを作成したところ、図7より発生箇所は40度付近、非発生箇所は20度付近という結果が得られ違いが見られた。そこで、処理を簡単にするために特徴量としてヒストグラムをベクトル化したものを使用した。ヒストグラムの階級数を5としているため、一つのヒストグラムで18次元得ることになる。傾斜角は全部で81個あるため一つの画像に対してヒストグラムは9つ作成され、傾斜角は全部で162次元となる(図8)。

(3) データ数

- ・土砂災害発生箇所：100枚
- ・土砂災害非発生箇所：30枚

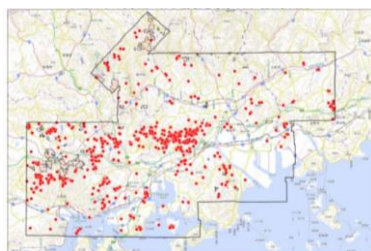


図3 広島県東部

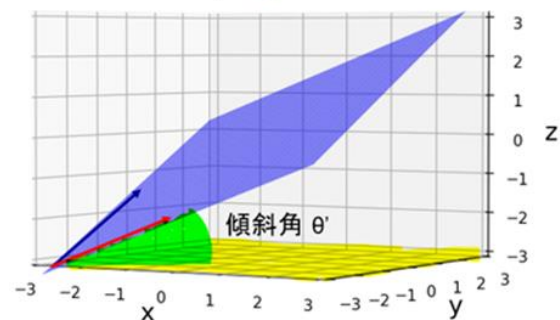
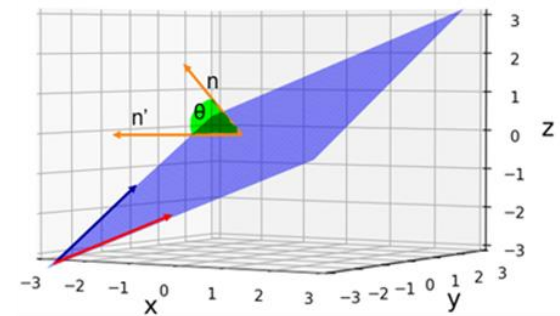
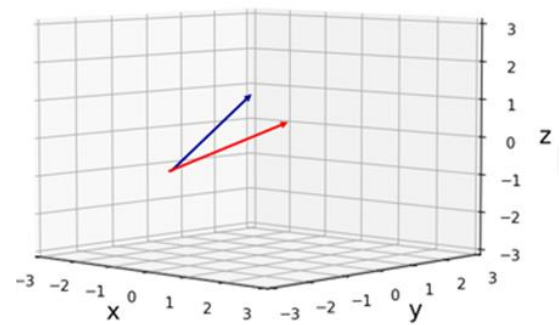


図6 傾斜角の算出方法

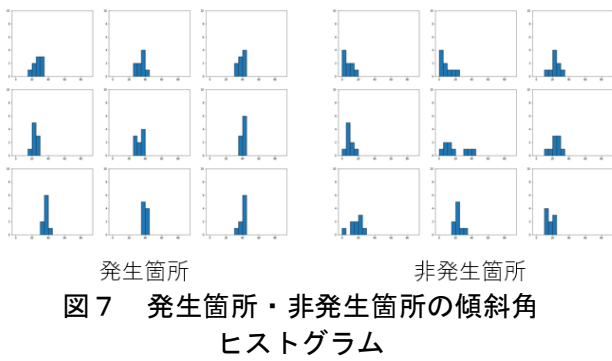
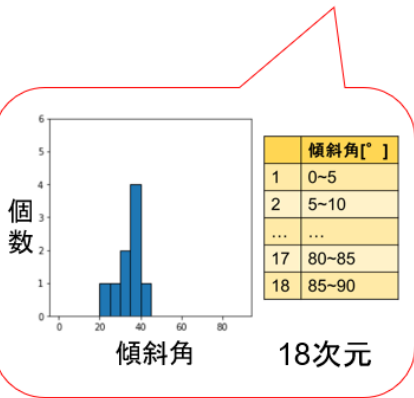
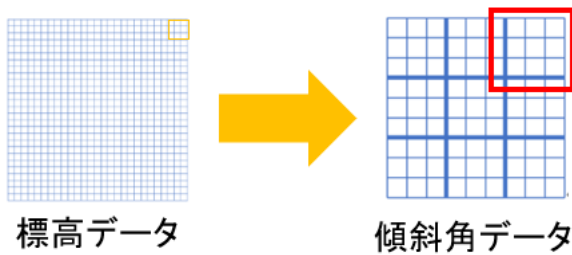


図 7 発生箇所・非発生箇所の傾斜角ヒストグラム



傾斜角は、全部で162次元(18×9)となる

図 8 特徴量の作成方法

4.3 教師なし外れ値検出によるモデル作成

学習には OCSVM[8], LOF[9], Isolation Forest[10] を用いた。機械学習には、Jupyter Notebook の Python3 を使用し、ライブラリである scikit-learn を使用した。

本研究は、教師なし外れ値検出でありハイパーパラメータの与え方により外れ値の検出精度が大きく変わることが知られている。そこで、最適なモデルを作成するために交差確認法とグリッドサーチを行った。表 1 に 3 つの手法で使用したハイパーパラメータを示す。

最適モデルを選択するための手順は以下のようになる。

- 100 個の異常データと 30 個の正常データを半分ずつパラメータ最適化用と実行用に分割する。分割は、重複なくランダムに抽出されるように行われる。

$$\text{異常データ} \begin{cases} 50(\text{最適化}) \\ 50(\text{実行用}) \end{cases} \quad \text{正常データ} \begin{cases} 15(\text{最適化}) \\ 15(\text{実行用}) \end{cases}$$

2. パラメータ最適化用データを用いてグリッドサーチによりハイパーパラメータの最適化を行う。
3. 2 で求めたパラメータを用いて実行用データで交差確認法を行い、精度を求める。
4. パラメータ最適化用と実行用のデータを変化させて同様に 10 回繰り返す。
5. パラメータの値を変更し、同様に繰り返す。

表 1 ハイパーパラメータ

アルゴリズム	ハイパーパラメータ
OCSVM [12]	nu : データが正常か異常かを判断する決定境界の単純さを表す gamma : 誤分類をどの程度許容するかを決める
LOF [13]	n_neighbors : 近傍数 contamination : データセット内の異常値の割合
ISF [14]	max_samples : Isolation tree を作成するために元のデータセットから選択するランダムサンプルの数 contamination : データセット内の異常値の割合

4.4 評価方法 1

作成したモデルは、識別結果の混同行列から一般的にモデル評価に用いられる Recall(再現率), Specificity(特異度), Precision(適合率)を算出して評価した。表 2 に本研究における混同行列の定義を示す。

表 2 混同行列

		予測	
		非発生	発生
実際	非発生	TN	FP
	発生	FN	TP

- Recall : 発生した箇所を正しく発生したと予測した割合

$$Recall = \frac{TP}{TP + FN}$$

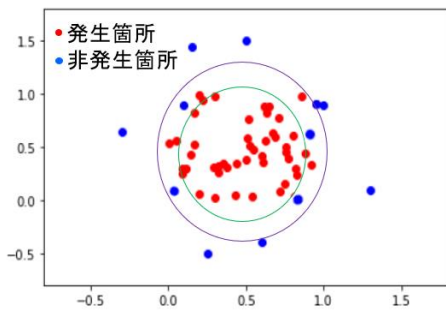
- Specificity : 発生していない箇所を正しく発生していないと予測した割合

$$Specificity = \frac{TN}{FP + TN}$$

- Precision : 発生したと予測した中で本当に発生していた割合

$$Precision = \frac{TP}{TP + FP}$$

本研究は、外れ値検出であるため性能を評価する際に閾値が重要となる。図 9 に示すように、閾値を緑とした場合と紫とした場合で再現率, 特異度, 適合率を算出した。閾値を小さくすると特異度が高くなり、閾値を大きくすると再現率が高くなっている。以上より、閾値と再現率, 特異度には関連性があるため再現率と特異度が重要な指標となっている。



➤ 緑の閾値

TP(●を●と予測)	40
FN(●を●と予測)	6
FP(●を●と予測)	0
TN(●を●と予測)	12
再現率	0.870
特異度	1
適合率	1

➤ 紫の閾値

TP(●を●と予測)	46
FN(●を●と予測)	0
FP(●を●と予測)	4
TN(●を●と予測)	8
再現率	1
特異度	0.667
適合率	0.92

図 9 閾値と再現率, 特異度の関係

5. 実験結果

実験結果を表 3 に示す。表 3 に示した結果は、4.3 でデータを 10 回変えて行った際に最も回数が多かったパラメータの平均値としている。

表 3 実験結果

	OCSVM	LOF	ISF
再現率	0.726	0.702	0.739
特異度	0.755	0.773	0.515
適合率	0.915	0.916	0.840
正解率	0.732	0.718	0.687

それぞれの差について t 検定を行ったところ有意水準を 0.05 とした場合、特異度と適合率において LOF と ISF, OCSVM と ISF の間に有意性があり、それ以外はなかった。

6. 考察

山斜面の傾斜角を用いることで土砂災害発生箇所を識別することができたが、なぜ可能なのかを調べるために傾斜角のデータに多次元尺度構成法[11]を実行し、土砂災害が発生したデータと発生していないデータの配置を可視化した。図 1 0 中の 1 は土砂災害が発生した箇所で、-1 が発生していない箇所を表している。図 1 0 より、1 が中央付近に-1 が端にあるのが見て取れる。これより、発生箇所と非発生箇所と分類できていることが分かる。そのため、識別することができたと考えられる。

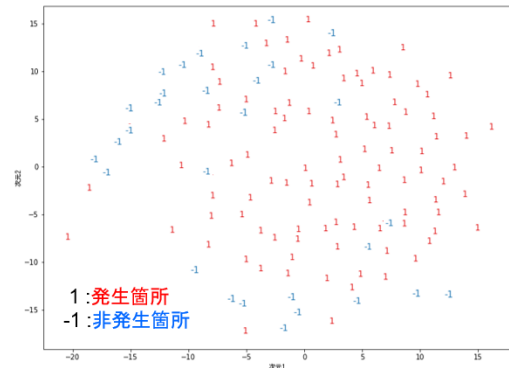


図 1 0 傾斜角における多次元尺度構成法

7. おわりに

本研究で使用した 3 手法の中では、OCSVM と LOF の精度が高いという結果が得られた。教師あり学習である先行研究と同等な再現率と特異度が得られたことから教師なし外れ値検出でも予測が可能であることが分かった。すなわち、土砂災害発生箇所のデータのみでの予測が可能であると考えられる。今回使用したデータは 2016 年のものであるが最近の環境変化に対応するためにも頻度が高いデータが好ましい。そこで、更新頻度が高い衛星から取得できるデータを用いての予測も行っていきたいと考えている。

参考文献

- [1] 内閣府, 防災情報のページ, 2017, <https://www.bousai.go.jp/kaigirep/hakusho/h18/bousai2006/html/honmon/hm01010101.htm>, (参照 2022-06-23)
- [2] 広島県庁, 土砂災害対策の現状とプランの基本方針, 2013, <https://www.pref.hiroshima.lg.jp/uploaded/attachment/128174.pdf>, (参照 2022-06-23)
- [3] 古川昭太, 丸山喜久, “共分散構造分析と機械学習に基づく地すべりと土砂崩壊地点の地形的特徴の評価”, 土木学会論文集, 74 巻, 4 号 (2018).
- [4] 石井唯嵩, 広兼道幸, 道財健斗, 倉本和正, “深層学習を用いた土砂災害警戒区域設定のためのデータセットの検討”, 土木学会論文集, 75 巻, 2 号 (2019).
- [5] 特定非営利活動法人, 土砂災害防止広報センター, <http://www.sabopc.or.jp/library/landslide/>, (参照 2022-06-23)
- [6] 国土交通省, 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>, (参照 2022-06-23)
- [7] 国土交通省, 国土地理院, 基盤地図情報サイト, <https://www.gsi.go.jp/kiban/>, (参照 2022-06-23)
- [8] Bernhard Scholkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, John Platt, “Support Vector Method for Novelty Detection”, Advances in neural information processing systems 12 (1999).
- [9] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, Jorg Sander, “LOF: Identifying Density-Based Local Outliers”, Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 93-104 (2000).
- [10] Liu, Fei Tony, Kai Ming Ting, Zhi-Hua Zhou, “Isolation forest”, 2008 eighth ieee international conference on data mining. IEEE, pp. 413-422 (2008).
- [11] Torgerson, Warren S, “Multidimensional scaling”, Theory and method, Psychometrika, 17.4, 401-419 (1952)
- [12] scikit learn, <https://scikit-learn.org/stable/index.html#>, (参照 2022-06-23)