

機械学習に基づく地震時の斜面崩壊予測

Prediction of slope failure during an earthquake based on machine learning

20WM1311 XIE HAOYANG

指導教員 丸山 喜久
劉 ウェン

SYNOPSIS

The 2018 Hokkaido Eastern Iwate earthquake caused a large number of slope failures and casualties in Atsuma town and its surrounding areas. In previous studies, several prediction methods for slope failures have been proposed, and the prediction accuracy has been discussed. However, the prediction models suitable for extensive areas are not solidly developed. In view of the current situation, this study uses machine learning techniques to predict the slope failures during the 2018 Hokkaido Eastern Iwate earthquake. In addition, the influence of NDVI of satellite image on the prediction results of machine learning model is discussed.

1. はじめに

2018 年北海道胆振東部地震では、厚真町を中心に数千箇所
の斜面崩壊が発生した¹⁾。斜面崩壊に対する事前対策を効率
的に行うには、斜面崩壊が発生する可能性の高い地点を事前
に把握することは重要である。

地震によって発生する斜面崩壊は、固有の地質と地形条件
に起因して発生することが多く、前兆的な地形に関しては発
生場所が予測可能とされている²⁾。Keefer³⁾は、ニューマーク
法⁴⁾に基づき危険斜面を抽出する方法を提案したが、地盤の
多様性の影響で広域的な予測の信頼性は高くない。中村ら⁵⁾
は、北海道胆振東部地震の際の厚真川流域の斜面崩壊予測式
を構築した。斜面勾配、平均曲率、地震動の最大加速度を説
明変数としており、斜面崩壊リスクマップの分類正誤率が
40.6%、空振り率が 81.4%とされている。

このように、広域に適用可能な斜面崩壊の予測手法は未だ
研究段階である。近年、急速に発展している機械学習を利用
して、広域に適用可能な斜面崩壊発生予測法を構築し、斜面
崩壊に関する警戒情報の精度向上を目指すことが期待されて
いる。また、リモートセンシング技術を利用した衛星画像は
広域性を有しており、面的な被害把握に非常に有用と考えら
れる。

そこで本研究では、2018 年北海道胆振東部地震を対象とし
て機械学習を用いた地震時の斜面崩壊推定に関する基礎検討
を行う。ロジスティック回帰分析、ランダムフォレスト、サ
ポートベクターマシンの 3 つの機械学習手法を用いて斜面崩
壊に関する要因を分析し、推定モデルを構築する。さらに、
地震後早期に衛星画像が取得できたものと仮定し、推定モデ
ルに厚真川流域の衛星画像の植生指標を考慮することを検討
する。最後に、対象地域における将来発生する斜面崩壊に備
えるために、推定モデルを用いて斜面崩壊の危険性を評価す
る。

2. 本研究で用いたデータ

本研究では、2018 年北海道胆振東部地震の際に丘陵や山地
で大規模な斜面崩壊が多発した震源に近い厚真町周辺を対象
領域とした(図-1)。国土地理院は、2018 年 9 月 6 日から 11
日にかけて撮影した空中写真から、「平成 30 年北海道胆振東
部地震に伴う斜面崩壊・堆積分布図」を作成している⁶⁾。本研
究では、これを 5 次メッシュ(250 m メッシュ)単位で集計し
た。斜面崩壊・堆積範囲を含む 250 m メッシュを「崩壊あり」、
含まないメッシュを「崩壊なし」とした。

梅田ら⁷⁾によると、崩壊地では顕著な湧水が確認できず、

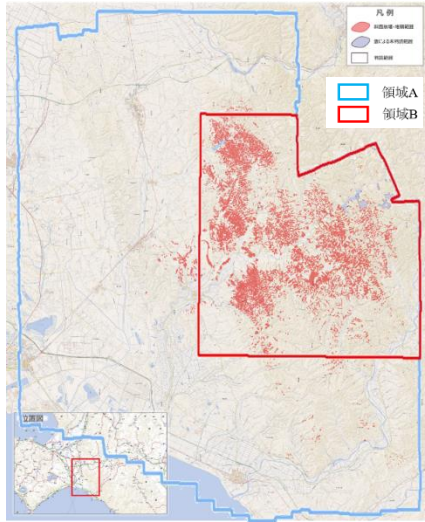


図-1 平成 30 年北海道胆振東部地震に伴う斜面崩壊・堆積分布図⁹⁾

表-1 設定領域の 5 次メッシュ数の内訳

	学習データ			テストデータ		
	崩壊あり	崩壊なし	合計	崩壊あり	崩壊なし	合計
領域A	2796	13271	16067	1219	5668	6887
領域B	2611	2246	4857	910	710	1620

主に地震時のせん断力によって崩壊が引き起こされたと考えられている。そこで、本研究は標高、傾斜角度、地震動の最大速度 (PGV)、微地形区分を説明変数とした斜面崩壊推定モデルの構築を目標とし、土壌の含水率を考えないこととした。

さらに、地震後早期に衛星画像が取得できたものと仮定し、正規化植生指数 (NDVI) を説明変数に加えた斜面崩壊推定モデルに関する検討も行った。

PGVは産業技術総合研究所が公開する地震マップ即時推定システム (QuiQuake) による推定分布を用いた。また、標高と傾斜角度は国土交通省が整備する国土数値情報から取得した。これらも、5 次メッシュ単位で集計されている。5 次メッシュごとの微地形区分のデータは J-SHIS 地震ハザードステーションから取得し、ダミー変数化して用いた。

本研究では、機械学習の適用範囲を、国土地理院の全判読領域 (図-1 の枠線内: 領域 A と領域 B) に設定した。設定領域の 5 次メッシュ数の内訳は表-1 の通りである。機械学習に適用する際、適用範囲が領域 A の場合、データの 75% を学習データ、25% をテストデータとした。適用範囲が領域 B の場合、データの 70% を学習データ、30% をテストデータとした。

3. 本研究で用いる精度評価指標

本研究では、機械学習による斜面崩壊の推定精度の評価に、正答率、適合率、再現率、F 値、マッシュズ相関係数、ROC 曲線、PR 曲線、 κ 係数の 8 つの評価指標を用いる。各指標式は式(2)~(6)の通りであり、式中の式の中の TP, FP, FN, TN は表-2 のように定義される。

表-2 精度評価のための混同行列

		予測結果	
		崩壊あり	崩壊なし
実際データ	崩壊あり	TP	FN
	崩壊なし	FP	TN

$$\text{正答率 (Accuracy)} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$\text{適合率 (Precision)} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{再現率 (Recall)} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F 値 (F-measure)} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

マッシュズ相関係数

$$(MCC) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(FN + TN)(TN + FP)(FP + TP)}} \quad (6)$$

また、崩壊ありと正しく判別する率を TPR (sensitivity), 崩壊なしを正しく判断する率を TNR (specificity) とする。

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

$$TNR = \frac{TN}{TN + FP} \quad (8)$$

斜面崩壊推定の際の閾値を変化させて TPR, TNR を算出し、横軸に 1-TNR, 縦軸に TPR をプロットしたものを ROC 曲線と呼ぶ。ROC 曲線の下部の面積 (Area Under the Curve) は AUC_{ROC} と呼ばれ、同様に、再現率を横軸に、適合率を縦軸にとった曲線を PR 曲線と呼ぶ。PR 曲線の AUC である AUC_{PR} は、クラス間のデータ数に偏りのある場合 (不均衡データ) の評価指標として有効である。

κ 係数の計算方法を式 (9) に示す。 p_0 は実際データと推定結果の実際の一致度、 p_e は実際のデータと推定結果が偶然一致する確率である。

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (9)$$

4. 領域 A の斜面崩壊推定

(1) ロジスティック回帰分析

ロジスティック回帰分析は、一つのカテゴリ変数 (二値変数) の確率を、複数の説明変数によって推定する多変量解析の一つである。 p は事象が発生確率、 β_0 は定数、 $\beta_1 \sim \beta_p$ は回帰係数、 $x_1 \sim x_p$ は説明変数である。発生確率 p を算出するためのロジスティック回帰モデルを式(10)に示す⁹⁾。

$$p(x) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)]} \quad (10)$$

領域 A を対象とした場合のロジスティック回帰分析のテストデータに対する推定結果と実際のデータを比較した結果を

図-2 に示す．非崩壊の F 値は高かったが，斜面崩壊の再現率は低かった．正答率や ROC 曲線による評価は良好であるが，MCC や κ 係数による実際のデータと推定結果の一致度は低かった．

ロジスティック回帰分析では，山地，丘陵，砂礫質台地，ローム台地のような斜面崩壊と関連の強い微地形区分に加えて，谷底低地の回帰係数が大きかった．領域 A の谷底低地の崩壊率は約 12.1% に達し，18 種類の微地形の中で 4 番目に大きい．このことが影響し，谷底低地の回帰係数が山地，丘陵等に次ぐ値となった．

(2) サポートベクターマシン

サポートベクターマシン(Support Vector Machine)では，あるパターン分類問題が線形分離不可能な場合，非線形写像によって特徴空間と呼ばれる高次元空間にマッピングする．式(11)で表されるヒンジ損失関数，式(12)で表される識別誤りを許容するソフトマージン方法を用いて分類問題を解く⁹⁾．ここで， ω は超平面の法線ベクトル， x_i は特徴ベクトル， b はスカラー変数， ζ_i はスラック変数である．

$$l_{\text{hinge}}(z) = \max(0, 1 - z) \quad (11)$$

$$\min_{\omega, b, \zeta_i} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \zeta_i$$

$$\text{s.t. } y_i(\omega^T x_i + b) \geq 1 - \zeta_i \quad \zeta_i \geq 0 \quad i = 1, 2, \dots, m \quad (12)$$

$$\kappa(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (13)$$

また，本研究では非線形写像に式(13)で表される Gauss カーネル関数を用いた．グリッドサーチに基づきハイパーパラメータ γ と C の最適値を定めた． γ の範囲を $10^{-5} \sim 10^5$ ， C の範囲を $10^{-2} \sim 10^2$ と設定して，5 分割の交差検証法によって適切なパラメータを求めた．

SVM の推定結果と実際のデータを比較した結果を図-2 に示す． AUC_{ROC} がロジスティック回帰分析よりも小さいものの，FP のメッシュ数は少ない．斜面崩壊の再現率は 0.66 となった．MCC は 0.6 を超え，良好な結果を示した．このことから，SVM の推定結果はロジスティック回帰分析より精度が高いことが分かる．

(3) ランダムフォレスト

ランダムフォレスト(Random Forest)は，多数の決定木を用いて識別性能の高い新たな識別器を構築するアルゴリズムである．決定木学習とは，学習データの特徴に応じて木構造の分類器を生成する学習手法であり，式(14)の情報利得(Information Gain)が最大となるようにする⁹⁾．

$$\text{Gain}(D, a) = \text{Ent}(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} \text{Ent}(D^v) \quad (14)$$

ここで， D は学習データ， a は特徴， D^v は学習データ D のうち特徴 a に対応する v 番目のサンプル， V は個々の木を分割するノード数である． $\text{Ent}(D)$ はエントロピーであり，式(15)のように書ける⁹⁾．

$$\text{Ent}(D) = - \sum_{k=1}^{|V|} p_k \log_2 p_k \quad (15)$$

$\text{Ent}(D)$ が小さいほど， D の純度が高い． p_k は学習データ D のうちクラス k の占める割合である．

本研究では，ランダムフォレストのハイパーパラメータの調整をグリッドサーチで行った．ハイパーパラメータの適切な範囲を設定して，5 分割の交差検証法によって適切なパラメータを求めた．また，RF のテストデータに対する推定結果から得られた各特徴量の影響度評価には，組み込み法によるモデルベース特徴量選択方法を使用した．領域 A では傾斜角度の影響度が最も高かった．標高と PGV の影響度は，それぞれ 0.288，0.243 と高く，これらも重要な説明変数と考えられる．

領域 A を対象とした場合について，テストデータに対する RF の推定結果と実際のデータを比較した結果を図-2 に示す．SVM の推定結果に比べると， AUC_{ROC} と AUC_{PR} の値が大きかった．MCC と κ 係数も RF の方が少し高く，それぞれ 0.62，0.616 であった．このように，総合的に評価すると，RF の推定精度が 3 つの手法の中で最も高かった．

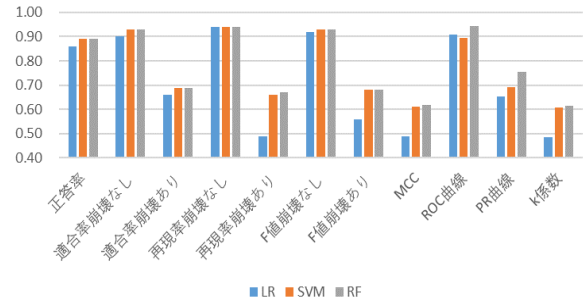


図-2 領域 A における精度評価指標の比較

5. 領域 B の斜面崩壊推定

(1) ロジスティック回帰分析

ロジスティック回帰分析の結果を図-3 に示す．説明変数に NDVI データを含めると正答率が 11% 向上し，FP と FN の数が 185 メッシュ減少した．全ての精度評価指標が NDVI を含めたモデルの方が高かった．しかしながら，NDVI を説明変数に加えることによってロジスティック回帰分析の推定精度は向上するものの， κ 係数は 0.581 にとどまり，推定結果と実際のデータの一致度は良好とは判断できなかった．

(2) サポートベクターマシン

SVMの結果を図-3に示す。精度評価指標に関しては、斜面崩壊の再現率を除き、全ての指標が説明変数にNDVIを含めたモデルの方が大きかった。特に、 AUC_{ROC} と AUC_{PR} は0.911, 0.93となり、判別能力が十分高いモデルと判断できる。混同行列を見ると、NDVIを含めた効果は非崩壊の識別に強く反映され、説明変数にNDVIを含めない場合よりもTNが124メッシュ増加した。

(3) ランダムフォレスト

RFの結果を図-3、図-4に示す。全ての精度評価指標が、説明変数にNDVIを含めたモデルの方が大きかった。MCCは0.73、非崩壊のF値は0.85、斜面崩壊のF値は0.88となった。

精度評価指標の比較を図-3に示す。MCCとF値を比較すると、RFが最も高かった。混同行列を見ると、説明変数にNDVIを含めた効果は、SVMと同様に、非崩壊の識別に反映された。また、NDVIの影響度は0.518となり、斜面崩壊を推定するのに最も重要な説明変数となった。なお、NDVIの値は微地形区分や傾斜角との相関性がある可能性もあるため、斜面崩壊との因果関係について今後詳細な検討を行う必要がある。領域BのテストデータでNDVIが0.0以下のメッシュを斜面崩壊の推定結果とし、実際データと比較した。RFの推定結果に比べてFPとFNの和が約2倍になった。また、正答率はRFの方が3%高く、MCCもRFの方が約0.2高かった。このように、機械学習手法を利用することで斜面崩壊の推定精度の向上が期待できる。

6. まとめ

本研究では、機械学習を利用して地震動強さ、微地形区分、標高、傾斜角度を考慮し、2018年北海道胆振東部地震における斜面崩壊の推定モデルを構築した。機械学習の手法には、ロジスティック回帰分析、RF、SVMの3手法を用いた。領域Aにおける正答率は、ロジスティック回帰分析が86%、SVMが89%、RFが89%となり、良好な精度が得られた。各モデルの精度評価指標を比較するとランダムフォレストの結果が最も良好であることが分かった。

さらに、地震後早期に衛星画像が取得できたことを仮定し、広域の斜面崩壊推定に衛星画像を活用するために正規化植生指数(NDVI)を機械学習の説明変数として用いた。NDVIを説明変数に含めた場合、ロジスティック回帰分析、SVM、RFの3つのモデルで推定結果と実際のデータの一致度が高くなった。特にSVMとRFで、 κ 係数がそれぞれ0.673、0.727となり、一致度が十分高いと判断された。このように、NDVIは斜面崩壊の推定精度の向上に大きく貢献した。

本研究が対象とした北海道胆振東部地震の被災地域の地形

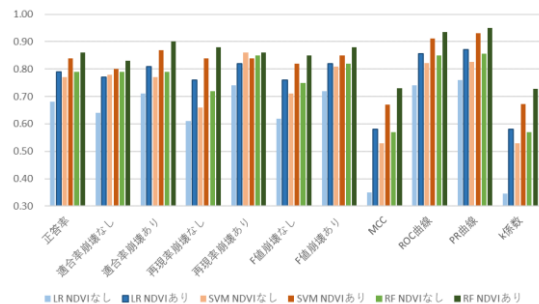


図-3 領域Bにおける精度評価指標の比較

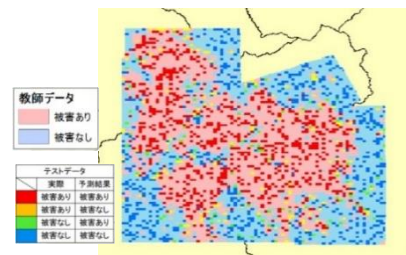


図-4 領域BにNDVIを含めたランダムフォレスト斜面崩壊評価図

は、全体的に低平であり大部分は標高170m～400mの山地・丘陵である。また、6000箇所の土層すべりが発生するなど、既往地震と比べて斜面崩壊が多数生じた²³⁾。このため、本研究で構築した推定モデルが他の地震に適用可能かは疑問が残る。そのため、今後は斜面崩壊の地域性やNDVIの季節性と地域性変化に注目し、他の地震で発生した斜面崩壊のデータを使って、本研究の機械学習モデルの汎用性を高める予定である。

参考文献

- 1) 村上泰啓, 水垣滋, 藤波武史: 平成30年北海道胆振東部地震における総崩壊地箇所数及び面積について, 令和2年度砂防学会北海道支部研究発表会, 2p, 2020.
- 2) 千木良雅弘: 地すべり・崩壊の発生場所予測—地質と地形からみた技術の現状と今後の展開—, 土木学会論文集C, Vol. 62, No. 4, pp. 722-735, 2006.
- 3) Keefer, D.: Landslides caused by earthquakes, Geological Society of America Bulletin, Vol. 95, pp. 406-421, 1984.
- 4) Newmark, N. M.: Effects of earthquakes on dams and embankments, Geotechnique, Vol. 15, pp. 139-160, 1965.
- 5) 中村香也, 真野拓司, 吉川契太郎: 厚真川流域を対象とした予測式による斜面崩壊リスクマップの作成について, 第64回(2020年度)北海道開発技術研究発表会論文, 6p, 2021.
- 6) 国土地理院: 平成30年(2018年)北海道胆振東部地震に関する情報, <https://www.gsi.go.jp/BOUSAI/H30-hokkaido-iburi-east-earthquake-index.html>
- 7) 梅田浩司, 千代谷佳織, 柴正敏, 近藤美左紀: 平成30年北海道胆振東部地震に伴う斜面崩壊, 応用地質, 第60巻, 第4号, pp. 172-179, 2019.
- 8) Zhou, Z.H.: Machine Learning, Springer Nature Singapore Pte Ltd., 2021.