斜面崩壊発生予測モデルの構築における機械学習のアルゴリズムの考察

鹿島建設　正会員　○越村　謙正
鹿児島大学　正会員　伊藤　真一
大阪産業大学　正会員　小田　和広

1. はじめに

筆者らは，豪雨による斜面崩壊の発生予測モデル（以下，モデルと呼ぶ）による土砂災害の危険度評価技術に関する研究を行っている．先行研究[1)]では，約 1km 四方の地域メッシュ（以下，メッシュと呼ぶ）を単位領域とし，各メッシュの地形特性を機械学習によって考慮することで高い予測精度を持つモデルを構築した．本研究では，モデルの精度向上を目指し，異なる機械学習のアルゴリズムを用いてモデルを構築する．そして，各モデルの予測精度を比較することにより，斜面崩壊の発生予測モデル構築に対する有効な機械学習のアルゴリズムを考察することを目的とする．

2. 機械学習に適用するデータベース

本研究では，兵庫県丹波市周辺の約 20km 四方の地域メッシュを対象に，地域内の 400 個のメッシュについて，各メッシュの土砂災害に関わる特徴を表現するデータベース[1)]を機械学習に適用する．**表-1** は，機械学習に適用するデータベースの一部を示している．対象地域では，平成 26 年 8 月豪雨に伴う斜面崩壊が多数発生しており，400 個の内 65 個のメッシュ内で斜面崩壊の発生が確認された．これら 65 個のメッシュを崩壊メッシュと呼ぶ．このデータベースにより表現される斜面崩壊のパターンを学習することでモデルを構築する．

表-1　各メッシュの特徴を表現するデータベース[1)]

メッシュ番号	地形特性				降雨条件		斜面崩壊
	傾斜		ラプラシアン		降雨指数		
	平均 (°)	標準偏差 (°)	平均 (/m)	標準偏差 (/m)	60分間積算雨量 (mm)	土壌雨量指数	発生 or 非発生
52356000	18.69	15.37	5.52×10^{-5}	2.80×10^{-2}	18	173.1	非発生
52356001	29.03	15.12	-6.37×10^{-4}	4.06×10^{-2}	25	192.2	非発生
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
52357198	9.65	9.33	1.78×10^{-4}	2.28×10^{-2}	55	236.1	非発生
52357199	25.49	12.86	-8.33×10^{-4}	3.98×10^{-2}	48	220.1	非発生

3. 機械学習のアルゴリズム

機械学習により構築されるモデルは，識別器とも呼ばれ，入力したデータを識別し，それに紐付けられた値を出力することで予測が可能となる．本研究では，このデータの識別規則の構成法（アルゴリズム）の違いが，構築されるモデルの精度に与える影響を検討する．代表的な識別規則は，(a) 事後確率による方法，(b) 距離による方法，(c) 関数値による方法，(d) 決定木による方法に大別される[2)]．本研究では，それぞれの代表的なアルゴリズムである，(a) ナイーブベイズ（Naive Bayes；以下，NB と呼ぶ），(b) k 最近傍法（k-Nearest Neighbor 法；以下，kNN と呼ぶ），(c) サポートベクトルマシン（Support Vector Machine；以下 SVM と呼ぶ），(d) ランダムフォレスト（Random Forests；以下 RF と呼ぶ）[2)]用い，構築されたモデルの精度を比較することで各アルゴリズムの適用性を検討する．また，解析ツールとして，統計解析向けのプログラミング言語である R 言語の caret パッケージを使用した．

4. モデルの精度の評価方法

各アルゴリズムにより構築されたモデルの精度の評価方法として，400 個のメッシュからなるデータベースの内，300 個分を機械学習に適用するデータ（以下，学習データと呼ぶ），残り 100 個分を機械学習に適用しないデータ（以下，検証データと呼ぶ）に分け，学習データで構築されたモデルに学習データを入力することで学習精度を，検証データを適用することで予測精度を評価する．なお，ここでは，学習・検証データ間でデータ内に占める崩壊メッシュの割合が偏らないよう，崩壊メッシュ 65 個の内，学習データに 49 個，検証データに 16 個が含まれるよう分割を行う．この分割をランダムに 10 ケース行い，各ケースで構築されたモデルの，学習データと検証データに対する正確率に基づき精度評価を行う．加えて，未知の斜面崩壊の発生に対する予測精度を評価するため，検証データの崩壊メッシュ 16 個のデータに対する正確率についても検討する．

キーワード　斜面崩壊，機械学習
連絡先　〒574-8530　大阪府大東市中垣内 3-1-1　大阪産業大学工学部都市創造工学科　TEL: 072-875-3001

表-2 SVMにより構築されたモデルの精度

SVM	正確率（%）									
ケース	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
学習データ	95.67	94.67	94.33	94.33	95.00	95.67	95.00	96.00	94.33	96.00
検証データ	96.00	94.00	92.00	92.00	93.00	95.00	93.00	89.00	93.00	93.00
検証データの崩壊メッシュ	87.50	93.75	68.75	62.50	81.25	87.50	68.75	75.00	75.00	68.75

表-3 各アルゴリズムのモデルの正確率

アルゴリズム	各データに対する正確率の10ケースの平均（%）		
	学習データ	検証データ	検証データの崩壊メッシュ
NB	93.23	90.60	76.88
kNN	94.77	93.00	72.50
SVM	95.10	93.00	76.88
RF	100	92.50	76.88

5. モデルの精度の評価結果と各アルゴリズムの適用性の検討

表-2 は，結果の一例として，SVMにより構築されたモデルについて，学習データ，検証データ，そして検証データの崩壊メッシュに対する正確率を示している．学習データと検証データに対する正確率を比較すると，各ケースで，それら値の隔たりは小さい．しかし，検証データの崩壊メッシュに対する正確率に着目すると，ケース②では学習データと検証データに対する正確率との隔たりが小さいモデルが構築されている一方で，それ以外のケースでは，その隔たりが大きいモデルが構築されていることが確認できる．この傾向は，他のアルゴリズムで構築されたモデルでも確認されている．これは，ランダムに行ったデータ分割の影響によるものと考えられる．そこで，10ケースの評価結果を平均することで，データベースの学習データと検証データへの分割され方が，構築されるモデルの精度に与える影響を軽減でき，各アルゴリズムのモデル構築に対する適用性をより適切に評価することが可能になると考えた．

表-3 は，10ケース毎に各アルゴリズムで構築されたモデルの正確率の平均を示している．検証データならびに検証データの崩壊メッシュに対する正確率の平均を比較すると，SVM では比較的高い予測精度を持つモデルを構築できることが示唆された．また，他のアルゴリズムでも，一定の予測精度を持ったモデルの構築が可能であることが分かった．しかし，学習・検証データへの分割がモデルの予測精度に与える影響が示唆されるため，複数の分割ケースによるモデル構築を行った上でモデルを選定することが望ましいと考えられる．

学習データに対する正確率の平均を比較すると，RFでは正確率が100%となり，他のアルゴリズムよりも明らかに高い．しかし，検証データに対する正確率との差が，他よりも大きいことから，過学習 [2)]の傾向が示唆される．**表-1** に示すよう，本研究で使用したデータは，400個の7次元のベクトルデータであるが，対象地域の拡大や，各メッシュの特徴を表すパラメータの追加によるデータの増大が見込まれる．すなわち，本研究で使用したデータについては，上述のように，一定の予測精度を持ったモデルの構築が可能であったが，より，多くのデータを取り扱う場合には，過学習によりモデルの精度が低下することが懸念される．そのため，他のアルゴリズムによるモデル構築が望ましいと考えられる．

6. まとめ

本研究では，機械学習による斜面崩壊の発生予測モデル構築における，異なる4種類のアルゴリズムの適用性を検討した．本研究から得られた知見は以下の通りである．

1) 学習・検証データへの分割のされ方が異なるケース間では，検証データの崩壊メッシュに対する正確率が大きく異なることが示唆された．そのため，データ分割を10ケース行い，各ケースで構築されたモデルの正確率を平均し，比較することで各アルゴリズムのモデル構築における適用性を検討することができた．
2) いずれのアルゴリズムでも一定の予測精度を持ったモデルの構築が可能であった．また，SVMでは比較的高い精度を持ったモデルの構築が可能であることから，その適用性の高さが示された一方で，RF では過学習の傾向が見られ，他のアルゴリズムを用いたモデル構築が望ましいことが分かった．

参考文献

1) 越村謙正，小田和広，伊藤真一：斜面崩壊発生予測モデルの構築における地形特性パラメータの適用性，第53回地盤工学研究発表会，2018.
2) 平井有三：はじめてのパターン認識，森北出版株式会社，pp.1-70，pp.114-134，pp.175-197，2016.