# 共分散構造分析と機械学習に基づく 地すべり地形の特徴の評価

古川 昭太1・丸山 喜久2

<sup>1</sup>学生会員 千葉大学博士前期課程 大学院工学研究科 (〒263-8522 千葉県千葉市稲毛区弥生町1-33) E-mail: acua2168@chiba-u.jp

<sup>2</sup>正会員 千葉大学准教授 大学院工学研究科 (〒263-8522 千葉県千葉市稲毛区弥生町1-33) E-mail: ymaruyam@tu.chiba-u.ac.jp

本研究では国土数値情報と防災科学技術研究所の地すべり地形データベースなどを利用し,宮城県仙台市 における地形的特徴と地すべり発生の因果関係を,共分散構造分析と機械学習による分析を用いて推定し た.それぞれの手法について被害推定式を構築し,地すべり地形評価図として既存の地すべり地形分布図 との比較考察を行った.共分散構造分析では,地すべり発生の危険が高い地点が本来の地すべり発生地点 に比べて過剰に抽出された.機械学習については,サポートベクターマシン(SVM)とランダムフォレスト を適用後,オーバーサンプリングすることにより,地すべり発生地点の見落としを減らすことができた.

Key Words : landslide, Covariance Structure Analysis, Support vector machine, Random Forest

#### 1. はじめに

わが国では、自然災害の発生に伴って多くの地すべり の被害を被ってきた.2004年10月23日に発生した新潟県 中越地震では地すべり131ヶ所、崖崩れ115ヵ所、土石流 21ヵ所と多くの土砂災害の被害が出た<sup>1)</sup>.また2011年の 東北地方太平洋沖地震では北海道と青森を除く東北地方 と群馬・栃木・茨城・埼玉・群馬で大規模な範囲での斜 面崩壊が確認された<sup>2)</sup>.その他、広島県広島市北部で発 生した平成26年8月豪雨<sup>3</sup>を筆頭に、豪雨による土砂災害 も多く発生している.地震調査委員会では今後30年間で マグニチュード7クラスの地震が首都圏で起こる確率は 70%としており、大規模な土砂災害の発生も懸念される. このことから、どのような地形的特性が地すべりの発生 の原因になり得るのかをあらかじめ予測しておき、地す べりの危険性が高い地域を予測することは有意義である と考えられる.

そこで、本研究では複数の構成概念間の関係を検討で きる統計的手法の一つである共分散構造分析(Covariance Structure Analysis)<sup>4</sup>と、機械学習の分析手法であるサポー トベクターマシン(Support vector machine)<sup>5</sup>およびランダム フォレスト(Random Forest)<sup>6</sup>を用いて、過去に地すべりが 発生した地点において、標高や微地形区分などの地形的 特徴が地すべり発生にどのような影響を与えるのかを評 価するモデルをそれぞれ構築する.そして、分析手法ご との結果と既存の地すべり地形分布図と比較することで、 今後どのような地域で地すべり発生が起こり得るのかを 予測する.

## 2. 対象地域と使用データ

本研究の対象地域は宮城県仙台市とした.使用したデ ータは国土交通省が整備する国土数値情報<sup>7</sup>,微地形区 分<sup>8</sup>,防災科研の地すべり地形分布図<sup>9</sup>である.国土数値 情報は標高・傾斜角などの地形データが5次メッシュご とに格納されている.微地形区分は、既往研究を参考に した.地すべり地形分布図<sup>9</sup>は、地すべり地形を航空写 真から判読したものである.カテゴリ毎のメッシュ数を 表-1に、標高および地すべり地形分布をそれぞれ図-1, 図-2に示す.

# 3. 共分散構造分析に基づく地すべり発生確率の評価

共分散構造分析では,素因(観測変数)が誘因(潜在変数)の影響を受けるものとし,さらに,それらがいくつ

地すべり頻度 地すべり頻度 93 地すべり頻度 \_\_\_\_\_\_\_ (上~未満:m) ~100 い 類度 頻度 4795 2126 地すべり頻度 御測変数 観測変勢 観測変調 <u>頻度</u> 4556 <u>時代第3</u> (以上・ 0~3 3~6 稲度 1442 284 CI F 未通 地すべり頻度 11.11 7383 山麓地 306 7855 100~200 306 102 1091 丘陵 200~300 300~400 6~9 9~12 2608 火山地 2396 517 359 2548 77: 火山北 火山山麓地 火山性丘陵 1635 645 2633 640 1048 651 400~500 12~15 2215 1382 594 680 102 24 500~600 600~700 1318 536 15~18 1704 648 岩石台地 化百 18~21 551 2342 1274 489 1326 砂罐管台地 197 擅 60 編員 10 兆 ローム台地 谷底低地 700~800 1208 468 21~24 1177 479 354 27 斜 高 800~900 900~1000 991 375  $24 \sim 27$ 1351 450 824 56 甶 局状地 自然堤防 後背湿地 771 27~30 550 270 1328 363 微 1000~1100 1100~1200 513 134 30~33 962 181 419 0 33~36 71 316 67 557 地 415 0 1200~1300 1300~1400 1400~1500 36~39 39~42 旧河道 三角州•海岸低地 128 45 166 14 5 12 0 形 1114 59 11 33 0 3 42~45 砂州・砂礫洲 289 n 2 45~48 0 砂丘 0 0 53 2515 1978 559 平坦 北 0 612 田 その他の農 58 38 ₩115-13 11-13 〒115-13 ・砂州間低地 79 0 0 傾北東 森林 荒地 3160 712 15212 4294 埋立地 74 0 確 - 岩破 孫• 岩破 河原 3279 289 607 21 14 東南東 斜 +建物用地 3439 601 3401 42 67 o 方懾 2930 455 地 道路 河道 69 3 0 向南西西西 (2588) 鉄道 その他の用料 河川地及び); 33 727 湖沼 その他 2843 498 0 11 ō 利 2345 54 15 491 用 北西 2427 577 412 30 NBLL (以上~未満: 0~50 50~100 )<del>頻度</del> 515 海水域 海水域 ゴルフ場 35 126 新吉 10 J 0 10123 2013 6212 150 13 記 100~150 3577 1334 伏 2353 585 200~250 量 643 95 250~300 300~350 76 10 22991 4553 n





図-1 宮城現仙台市の標高データ

かの要因に集約できると仮定する.その要因と地すべり 発生の確率との相関を評価する.ここで、全ての観測変 数は平均0,標準偏差1に標準化してある.基本パスモデ ルを図-3に示す.なお、基本パスモデルは既往研究<sup>10</sup>を 参考に仮定した.





メッシュごとの地すべり発生確率をベイジアンモデル とファジーセットモデルの二つの手法により求める. な お,どちらも既往研究<sup>11)</sup>を参考にした.



図-2 宮城県仙台市の地すべり地形分布図

ベイジアンモデルでは、条件付確率を求める際に用いられるベイズの定理を利用することによって、地すべり発生確率を求める。メッシュ*q*において地すべりが発生すると想定される事象を{T<sub>q</sub>}と表し、メッシュ*q*に該当する「i番目の素因における*j*番目のカテゴリ」をC<sub>ij</sub>(*i*=1, …*m*,*i*=1,…*n*)と記述する.

ベイズの定理に従い、地すべり発生確率S<sub>q</sub>は以下のように記述できる.

$$S_q = \frac{P\{T_q\} \cdot P\{c_{1j}, c_{2j}, \dots, c_{mj} | T_q\}}{P\{c_{1j}, c_{2j}, \dots, c_{mj}\}}$$
(1)

ここで、各素因が独立事象であると仮定すると,

 $P\{c_{1j}, c_{2j} \dots, c_{mj} | T_q\} = P\{c_{1j} | T_q\} \dots P\{c_{mj} | T_q\}$ (2)

となり,式(2)を式(1)に代入すると,

$$S_q = \frac{P\{c_{1j}\}...P\{c_{mj}\}}{P\{c_{1j},c_{2j},...,c_{mj}\}} \cdot P\{T_q\} \cdot \frac{P\{T_q|c_{1j}\}}{P\{T_q\}} \cdots \frac{P\{T_q|c_{mj}\}}{P\{T_q\}}$$
(3)

となる. ここで, 式(3)を計算するためには,

$$P\{c_{ij}\} = \frac{N_{ij}}{A} \tag{4}$$

$$P\{c_{1j},\ldots,c_{mj}\} = \frac{N_X}{A} \tag{5}$$

$$P\{T_q\} = \frac{N_0}{A} \tag{6}$$

$$P\{T_q | c_{ij}\} = \frac{N_d}{N_{ij}} \tag{7}$$

を式(3)にそれぞれ代入すればよい.ここで、 $N_{ij}$ は i番 目の素因におけるカテゴリjのメッシュ数、Aは全メッ シュ数、 $N_x$ はメッシュqにおける $i \sim m$ 番目の素因に対 する全カテゴリにおいて等しい属性のメッシュ数、 $N_0$ は地すべり地形のメッシュ数、 $N_d$ はi番目の素因におけ るカテゴリjの領域に属する地すべり地形のメッシュ数 である.

ファジーセットモデルでは、ファジー理論を活用して 地すべり発生確率を求める.*X*を対象地域の全メッシュ の集合、あるメッシュに対する「地すべりが発生する可 能性」のファジー集合を*S、ファジーメンバーシップ*関 数をµ<sub>s</sub>とすると、ファジーメンバーシップ関数は以下の ように記述される.

$$\mu_{\rm s}: X \to [0,1] \tag{8}$$

**これは***X*上のファジー集合*S*が,区間[0,1]の間の値をとることを意味する.式(8)より,地すべり発生確率*S*<sub>q</sub>は以下のように記述できる.

$$S_q = \mu_s \{q | c_{1j}, \dots, c_{mj}\}$$
 (9)

式(9)を計算するために、ファジー演算の中でも多く使われているファジー代数和演算を採用すると以下の式に 帰着する.

$$S_q = \mu_s \{ q | c_{1j} \dots, c_{ij} \} = 1 - \prod_{i=1}^m \{ 1 - \frac{N_d}{N_{ij}} \}$$
(10)

ベイジアンモデル,ファジーセットモデルの両方で地 すべり発生確率を計算した後,図-3の基本パスモデルの 適合度を算出する.さらに,基本モデルから観測変数を 1または2個減らしたモデルの適合度を評価し,最適なパ スモデルを探索した.結果を表-2,表-3にまとめる.な

表-2適合度指標(ベイジアンモデル)

モデル名	削除した素因	RMSEA	CFI	TLI
モデルA	—	0.212	0.902	0.743
モデルB	標高	0.126	0.977	0.913
モデルC	起伏量、	0.158	0.955	0.83
モデルD	傾斜角	0.126	0.969	0.884
モデルE	微地形	0.164	0.966	0.873
モデルF	土地利用	0.259	0.908	0.654
モデルG	傾斜方向	0.298	0.904	0.639
モデルH	標高+起伏量	0.044	0.999	0.988
モデルΙ	標高+傾斜角	0.042	0.999	0.988
モデルJ	標高+微地形	0.084	0.997	0.972
モデルK	標高+土地利用	0.043	0.999	0.991
モデルL	標高+傾斜方向	0.242	0.979	0.787
モデルM	起伏量+傾斜角	0.09	0.996	0.957
モデルN	起伏量+微地形	0.061	0.998	0.981
<mark>モデルO</mark>	起伏量+土地利用	0.051	0.999	0.985
モデルP	起伏量+傾斜方向	0.31	0.96	0.596
モデルQ	傾斜角+微地形	0.071	0.997	0.974
モデルR	傾斜角+土地利用	0.056	0.99	0.981
モデルS	傾斜角+傾斜方向	0.238	0.975	0.753
<del>モデルT</del>	微地形+土地利用	0.013	1	0.999
モデルU	微地形+傾斜方向	0.318	0.968	0.682
モデルV	土地量+傾斜方向	0.515	0.908	0.083

表-13 適合度指標(ファジーセットモデル)

モデル名	削除した素因	RMSEA	CFI	TLI
モデルA	—	0.217	0.95	0.869
モデルB	標高	0.123	0.991	0.968
モデルC	起伏量	0.157	0.983	0.936
モデルD	傾斜角	0.155	0.984	0.939
モデルE	微地形	0.154	0.987	0.952
モデルF	土地利用	0.267	0.958	0.842
モデルG	傾斜方向	0.305	0.951	0.817
モデルH	標高+起伏量	0.045	1	0.996
モデルΙ	標高+傾斜角	0.042	1	0.997
モデルJ	標高+微地形	0.083	0.999	0.991
モデルK	標高+土地利用	0.04	1	0.997
モデルL	標高+傾斜方向	0.231	0.992	0.924
モデルM	起伏量+傾斜角	0.086	0.999	0.987
モデルN	起伏量+微地形	0.059	0.999	0.994
モデルΟ	起伏量+土地利用	0.048	1	0.996
モデルP	起伏量+傾斜方向	0.06	0.999	0.992
モデルQ	傾斜角+微地形	0.054	0.999	0.992
モデルR	傾斜角+土地利用	0.056	0.998	0.981
モデルS	傾斜角+傾斜方向	0.301	0.984	0.842
モデルT	微地形+土地利用	0.01	1	1
モデルU	微地形+傾斜方向	0.296	0.988	0.881
モデルV	土地量+傾斜方向	0.242	0.952	0.855

お、適合度の高いモデルは黄色で示す. RMSEAは共分 散構造分析において頻繁に利用されている適合度指標で あり、0以上の値をとり、値が小さいほどモデルの適合 が良いとされる. 広く用いられている判断基準では RMSEA>0.10ならば適合が悪く、RMSEA<0.05なら適合が 良いとされている<sup>10</sup>. CFIとTLIは独立モデルと解析され たモデルの適合度(カイ自乗値)を比較したモデルである. ここで、独立モデルとはあるデータの分析をする場合に、 想定しうる中で最も適合の悪いモデルのことである. 両 これらは、0~1の間の値をとり、1に近いほど適合が良 いと判断する<sup>13</sup>.



図-5 ファジーセットモデルにおけるパスモデル

表-2,表-3の結果から,分析手法によらず,適合度の 高いモデルは同じになった.これらのモデルでは,どの 適合度指標も一般的に良好とされている値を示しており, 妥当なモデルであると考えられる.そこで,これらの5 つのモデルを対象として,詳細な分析を行う.

これらの5つのモデルと基本モデルのパス係数(相関 係数)を図-4,図-5に示す.それぞれのパスモデルの相 関係数から,地すべりには要因1(標高,起伏量,傾斜 角,傾斜方向)が大きく影響しており,要因2(微地形 区分),要因3(土地利用)の影響は小さいということ がわかる.また、要因1内の相関係数を見ると、どのモ デルにおいても標高と傾斜角の相関係数が大きいことが わかる.反対に傾斜方向についてはどのモデルについて も相関が小さかった.

ここで、それぞれの地すべり地形の的中率を算出し、 最も的中率の高いモデルを採用することとする.具体的 には、横軸にパスモデルをもとに算出した標準化された 地すべり発生確率を、縦軸にその度数および累積頻度を 示す.さらに、負極側から地すべり地形の累積頻度曲線 を、正極側から非地すべり地形の累積頻度曲線を描く (図-6). つまり,地すべり地形をAグループ,非地すべり地形をBグループとすると,

$$F_{A}(X) = 1 - \int_{-\infty}^{X} f_{A}(x) dx \quad (\int_{-\infty}^{+\infty} f_{A}(x) dx = 1) \quad (11)$$

$$F_{\rm B}(X) = \int_{-\infty}^{X} f_{\rm B}(x) dx \qquad (\int_{-\infty}^{+\infty} f_{\rm B}(x) dx = 1) \quad (12)$$

となり、 $F_A(X)$ ,  $F_B(X)$ の交点が判別の閾値となる.

手法ごとのそれぞれのモデルについての的中率を調べたところ、ベイジアンモデルでは73.9%、ファジーセットモデルでは74.00%と、どちらの手法においても、基本モデルから起伏量と土地利用を抜いたモデルであるモデルOの的中率が最も高い結果を示した(図-7). したがって、本研究の共分散構造分析ではモデルOを採用することとする.

#### 4. 機械学習に基づく地すべり発生箇所の評価

本研究では、共分散構造分析に加えて、機械学習に基づく分析を実施した.ここでは、サポートベクターマシンとランダムフォレストの二手法を利用した.全データの10%を学習データとして地すべり発生箇所の分類器を作成し、残りの90%をテストデータとして分類機の精度評価に利用した.

#### (1)サポートベクターマシン

サポートベクターマシン(SVM)は、教師あり学習を用いたパターン認識手法の一つである. 基本的には2つの クラスを識別するための識別機を構成するための学習手 法とされており、認識性の優れた学習モデルの一つとい われている.

SVMによる2クラス分類では、式(13)で表される最適 化問題を解く.ここで、wは分離超平面の法線ベクトル、 ¢は入力ベクトルxを特徴空間Fへ非線形写像する関数





図-7 各モデルの的中率

(式(14)), bはスカラー変数を表す<sup>14</sup>.

$$\min_{\boldsymbol{w},\boldsymbol{b},\boldsymbol{\zeta}} \frac{1}{2} \| \mathbf{w} \|^{2} + C \sum_{i \in [n]} \zeta_{i}$$
s.t.  $y_{i}(\mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_{i}) + \boldsymbol{b}) \geq 1 - \zeta_{i}, i \in [n], \zeta_{i} \geq 0, i \in [n]$ 
(13)

$$K(\mathbf{x}_{i}, \mathbf{x}_{i}) = \boldsymbol{\phi}(\mathbf{x}_{i})^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_{i})$$
(14)

さらに、本研究では非線形写像のための関数にRBFカー ネルを用いる.

$$K(\mathbf{x}_{i}, \mathbf{x}_{j}) = \exp(-\gamma \|\mathbf{x}_{i} - \mathbf{x}_{j}\|^{2})$$
(15)

SVM		実際のデータ		
		危険あり	危険なし	
分析	危険あり	721	406	
結果	危険なし	3395	16170	



R F		実際のデータ		
		危険あり	危険なし	
分析	危険あり	1439	1050	
結 果	危険なし	2677	15526	

x 4 SVMおよびフラクムフォレスト(RF)による力が相来と実际の地方、り先主地点の

全体:81.63%,地すべり発生地点:17.52% (C=2.5 \* 10<sup>-1</sup>,γ=5.01 \* 10<sup>-2</sup>)



式(15)のγと式(13)のCは、ハイパーパラメータと呼ばれ 分類結果に大きく影響する<sup>14</sup>.

説明変数には、共分散構造分析において最も的中率の 高かったモデル Oと同じく,基本モデルから起伏量と 土地利用を除いた変数を利用した. また, グリッドサー チ 5に基づきハイパーパラメータを設定する. グリッド サーチとは、Cとyの値の範囲を任意に設定し、交差検 証法により、最適な C と γ の値を求める手法である.本 研究では、全データの10%を学習データ、γを105~105、 Cを10<sup>2</sup>~10<sup>2</sup>の範囲に設定して、交差検証法(k=10) によって最適なパラメータを求めた.地すべり有無の分 類結果、的中率(地すべりの有無を正しく分類した割合) と、地すべり地形のメッシュだけで算出した的中率、ハ イパーパラメータである Cと yの値を表-4に示す.的中 率は共分散構造分析に比べ向上したが、地すべり発生地 点の抽出精度が悪く、特に実際に地すべりの危険がある 地点を危険なしと判定する危険個所の見落としが多くみ られた.

#### (2) ランダムフォレスト

ランダムフォレスト (RF) とは、複数の決定木を用 いる集団学習法(アンサンブル学習)の一つである。各決 定木での予測結果を多数決することにより、結果の取得 を行う。決定木学習とは、データの種類に応じて決定木 を成長、分類させていく学習手法であり、式(16)に示す 情報利得 IG が最大となるようにする<sup>10</sup>.

$$IG(D_{p}, f) = I(D_{p}) - \sum_{j=1}^{m} \frac{N_{j}}{N_{p}} I(D_{j})$$
(16)

ここで、D<sub>p</sub>は学習データ、Nはノード、*j*は注目してい るデータ、*I*は不純度、*m*は特徴量である.不純度はデ ータに偏りがあるほど、大きな値になる指標であり、エ ントロピー、ジニ係数、分類誤差などが用いられる.式 (17)はエントロピーの式<sup>ID</sup>を表している.

 $I(N) = -\sum p(k|N)logp(k|N)$ (17)

ここでp(k|N)は、各ノードNでクラスkを取る確率である.

また,特徴量 m は個々の木を分割するノード数であ り, R 言語を用いたランダムフォレストでは, この特徴 量 m をハイパーパラメータとして分類器を構成してい くこととなる<sup>18</sup>. 説明変数は前章のモデル O と同じに した. 表-4 に分析結果および特徴量 mの値を示す. SVM による分析と同様に, 的中率そのものは向上したものの, 地すべり発生地点の抽出の精度が低く, 危険個所の見落 としが多くみられた.

#### (3)不均衡データの整形

SVMとRFの結果,地すべり発生地点の見落としが多かったことが課題として挙げられた.その原因としては、本研究で取り扱っている地すべり地形分布図のデータが、 正例(地すべり発生が認められるメッシュ)に対して負例 (地すべり発生が認められないメッシュ)の数が極端に多い不均衡データであるため、負例を過剰に抽出しまった ことが考えられる.そこで、本研究ではこのような不均 衡データにおける誤分類を解決するための手法として、 サンプリング法の一つである SMOTE (Synthetic Minority Over-sampling Technique)<sup>19</sup>を適用した.

SMOTEとは、Chawla らによって提案されたオーバー サンプリング法の一つであり、k-最近傍法を基にしたア ルゴリズムである.任意の正例を指定し、類似度計算に より最近傍にある他の正例を特定する.そして、その二 点を結ぶ直線上に新たな正例を作成する.この作業を指 定の回数繰り返すことで正例の個数を増やし、データ間 の偏りを減少させる<sup>20</sup>.式(18)は類似度計算の式である.

$$sim(m_a, m_i) = \sqrt{\sum_{j=1}^n (v_{a,j} - v_{i,j})^2}$$
 (18)

ここで,*m*は任意の点,vは点*m*における説明変数,*n* は説明変数の数を表す.

本研究では、正例と負例の比が1:4なので、データが 均衡になるように、正例を4倍にオーバーサンプリング

表-5	オーバーサン	/プリング後の	SVM および RF によ	る分析結果と実際の	)地すべり発生地点の比較
-----	--------	---------	---------------	-----------	--------------

	SAM		実際のデータ		
		V IVI	危険あり	危険なし	
	分析	危険あり	3306	5007	
	結 果	危険なし	810	11569	

全体:71.89%,地すべり発生地点:80.32% (C=64.00,y=3.1 \* 10<sup>-2</sup>)

R F		実際のデータ		
		危険あり	危険なし	
分析	危険あり	3164	4834	
結 果	危険なし	952	11742	

全体:72.04%,地すべり発生地点:76.87% (m=20)



図-8 地すべり地形評価図の比較

して学習データを作成,分析を行った.表-5に SMOTEによって不均衡データ整形後の SVM および RF の分類結果を示す.表-4と比べて,的中率は下がったが,地すべり発生地点の抽出の精度の向上がみられ,オーバーサンプリングの効果と考えられる.

#### 5. 地すべり地形評価図

共分散構造分析と機械学習での二手法の分析の結果と, 既存の地すべり地形分布図との比較を行い,地すべり地 形評価図を作成した(図-8).共分散構造分析の結果と しては,分析結果で危険とされた地点が実際の地すべり 発生地点よりも過剰に抽出されている.SMOTEによる オーバーサンプリングを適用していないSVMによる地 すべり地形評価図は,危険箇所の見落としがかなり多く なっていることがわかる.

SMOTEによるオーバーサンプリング適用後の SVM および RFの結果は、危険個所の見落としが減少し、共 分散構造分析と似た結果となった.このことから、危険 箇所の見落としは増やさずに、過剰に抽出されている地 点を減らしていくことが、今後の課題として挙げられる.

# 6. まとめと今後の課題

本研究では、共分散構造分析とサポートベクターマシン、ランダムフォレストの二種類の機械学習手法に基づき、宮城県仙台市の地すべりの危険箇所の特徴を評価した.共分散構造分析では地すべり発生には標高、傾斜角、

傾斜方向, 微地形の4つの素因が影響し, とくに標高と 傾斜角が大きく影響することがわかった. 機械学習によ る分析では, 共分散構造分析の結果よりも高い的中率を 得ることが出来たものの既存の地滑り危険箇所の見落と しが課題となった. そこで, オーバーサンプリング法の 一つである SMOTE を適用したところ, 地すべり箇所の 見落としを減少させることができた.

今後は、さらなる素因の選定を行うとともに、各手法 のハイパーパラメータの設定や、オーバーサンプリング の調整、別の分析手法に関する検討を行う.

#### 参考文献

- 新潟県土木部砂防課:新潟県中越地震と土砂災害, http://www.pref.niigata.lg.jp/HTML\_Article/864/780/chuetsujisin.pdf
- 2) 国土交通省:災害・防災状況,http://www.mlit.go.jp/saigai/ saigai\_110311.html
- 国土交通省砂防部:平成26年8月豪雨による広島県で発生した土砂災害への対応状況,http://www.mlit.go.jp/river/sabo/H26\_hiroshima/141031\_hiroshimadosekiryu.pdf
- 4) 豊田秀樹: 共分散構造分析[R編], 東京図書, 2014.
- 5) 栗田多喜夫:サポートベクターマシン入門, http://
- home.hiroshimau.ac.jp/tkurita/lecture/svm.pdf
- 6) 波部斉:ランダムフォレスト, https://www.slideshare. net/HitoshiHabe/ss-58784421
- 7) 国土交通省:国土数値情報ダウンロードサービス, http://nlftp.mlit.go.jp/ksj/
- 8) 若松加寿江,松岡昌志,久保純子,長谷川浩一,杉浦正 美:日本全国地形・地盤分類メッシュマップの構築,土 木学会論文集,No.759/1-67,213-232,2004.

- 防災科学技術研究所:地すべり地形分布図データベース,15) 荒川正幹,宮尾知幸,船津公人:ドラッグライクネスモデル 9) の構築とその可視化, Journal of Computer Aided Chemistry, Vol.9, http://lsweb1.ess.bosai.go.jp/ 10) 小島尚人,大林成行,青木太:共分散構造分析を導入し pp. 70-80, 2008. た斜面崩壊危険箇所評価アルゴリズムの構築,土木学会 16) 波部斉: ランダムフォレスト, コンピュータビジョンとイメ ージメディア, pp. 1-8, 2012. 論文集,No.714/VI-56,pp.79-93,2002. 11) 大林成行,小島尚人, Chang-Jo F.Chung:斜面安定性評価モ 17) 山岡 啓介:ランダムフォレスト,映像情報メディア学会誌, デルの精度比較とその実用化への提案,土木学会論文集, Vol. 66, No. 7, pp. 573-575, 2012. 18) 小林寛武,, 戸田航史, 亀井靖高, 門田暁人, 峯恒憲, 鵜林 No.630/VI-44, pp.77-89,1999. 12) 星野崇宏,岡田謙介,前田忠彦:構造方程式モデリング 尚靖:11 種類の fault 密度予測モデルの実証的評価,電子情 における適合度指標とモデル改善について:展望とシミュ 報通信学会論文誌 D, Vol.96, No. 8, pp. 1892-1902, 2013. Rpubs:SMOTE で不均衡データの分類, https://rpubs.com/ レーション研究による新たな知見, 行動計量学, 第 32 巻, 19) 第2号, pp.209-235, 2005 hoxo m/54954 13) 廣野元久: グラフィカルモデリングのための G-GM & L-20) 亀井靖高,門田暁人,松本健一:Fault-proneness モデルへのオ GMデータ解析システム,計算機統計学, Vol. 15, No. 1, pp. ーバーサンプリング法の適用, http://www.empirical.
  - jp/paperdb/papers/archive/102/102.pdf マシンン 供給抽答型プ
- 14) 竹内一郎, 鳥山昌幸:サポートベクトルマシン(機械学習プロフェッショナルシリーズ),講談社, 2015

63-74,2003.

EVALUATION OF THE TOPOGRAPHIC CHARACTERISTICS AT THE LANDSLIDE OCCURENCES BASED ON THE COVARIANCE STRUCTURE ANALYSIS AND THE MACHINE LEARNING

### Shota Furukawa and Yoshihisa MARUYAMA

This study evaluated the topographic characteristics at the landslide outbreak locations based on the covariance structure analysis and the machine learnings. The support vector machine (SVM) and random forest (RF) were employed in this study to achieve the objective. The estimated results were compared to investigate the accuracy of the detection of landslide occurrences. The SVM and RF gave reasonable results if a technique for oversampling is applied to correct for a bias in the training dataset. Lastly, the discriminated results were projected onto a map to evaluate the possibility of landslide occurrences.