

共分散構造分析と機械学習に基づく液状化発生地点の予測

Estimation of liquefaction occurrences based on covariance structure analysis and machine learning

17WM1342 柳瀬 匡雄
Masao Yanase

指導教員 丸山 喜久

SYNOPSIS

This study tries to develop a method to estimate the locations of soil liquefaction occurrences based on the covariance structure analysis, support vector machine (SVM) and random forest (RF). The risk assessment map for soil liquefaction is illustrated, and it is compared with an existing map to show liquefaction evidences. To achieve the objective, the datasets of liquefaction occurrences after the 2011 off the Pacific Coast of Tohoku earthquake and the 2016 Kumamoto earthquake were employed in this study. The estimated results of liquefaction occurrences in this study were more accurate than that of the previous study.

1. はじめに

2011年東北地方太平洋沖地震では、青森県から神奈川県までの南北方向に約650kmの広域な範囲で液状化が発生したり。液状化は、建物や橋梁などの様々な構造物やライフライン施設等に甚大な被害を与えることがある。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を70%と評価しており、首都圏の液状化対策は一層重要である。このことから、将来起こりうる地震に対して液状化が発生する危険がある地点を予測し、対策を講じておくことは意義があるものと考えられる。

広域の液状化危険度予測を行う場合には、まず地形、地質や液状化履歴をもとにした概略の予測方法によって危険性を有する地域を絞り込み、その中で必要に応じて一般の土質調査による簡易の予測方法あるいは詳細な予測方法を選択する手段が採られる。しかし、近年首都圏をはじめとして、大都市域では都市相互の関係は非常に密接化しているため、行政区域にとらわれず、さらに広域の都市群に対して液状化危険度を総合的に評価することが望まれる。

そこで本研究では、まず複数の構成概念間の関係を検討できる統計的手法である共分散構造分析 (Covariance structure analysis)²⁾を行い、どの素因が液状化の発生に大きく影響するのかを明らかにする。さらに、教師あり学習に基づくパターン認識手法の一つであるサポートベクターマシン(Support vector machine)³⁾と複数の決定木を用いる集団学習法(アンサンブル学習)の一つであるランダムフォレスト(Random forest)⁴⁾に基づく機械学習を実行し、比較的少数の説明変数で液状化発生地点を予測する手法を構築することを目的とする。

2. 対象地域と使用データ

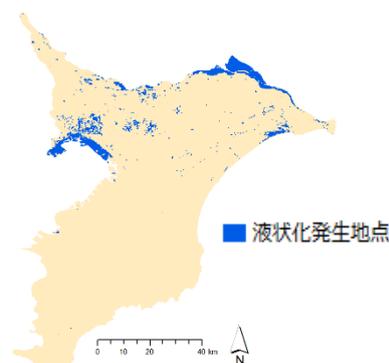
東北地方太平洋沖地震の際の千葉県での液状化発生地点を対象に分析を行った。使用したデータは国土交通省が整

備する国土数値情報⁵⁾と、微地形区分⁶⁾である。前者は標高、傾斜角などの地形データが5次メッシュ(約250mメッシュ)ごとに整備されている。後者には微地形が5次メッシュごとに格納されている。本研究で用いる地形的特徴データは、標高、傾斜方向、最大傾斜角、起伏量(標高の最大値と最小値の差)、微地形区分、河川からの距離とし、さらに東北地方太平洋沖地震における計測震度⁷⁾を地震動指標に用いた。これらの説明変数の選定は、松岡ら⁸⁾による液状化危険度の推定手法等を参考にした。

東北地方太平洋沖地震による液状化発生地点は、若松ら⁹⁾がとりまとめたデータ¹⁾を用いた。液状化発生地点が5次メッシュごとに格納されていたため、解析のメッシュサイズを5次メッシュとした。なお、本研究で使用した千葉県のメッシュ総数は79695メッシュであり、そのうち3702メッシュが液状化発生地点である(図-1)。

3. 共分散構造分析に基づく液状化発生確率

共分散構造分析では、誘因が各素因に影響し液状化を引き起こすものと仮定し、素因をいくつかの要因にまとめることができるものとして、要因と液状化発生確率との相関

図-1 千葉県の液状化発生地点¹⁾

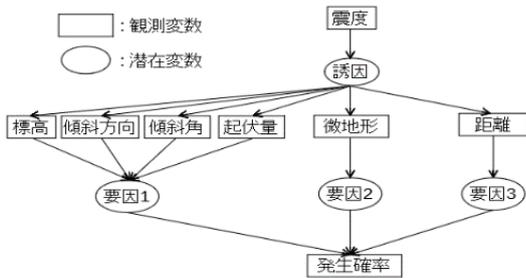


図-2 本研究で仮定した液状化発生の基本モデル

を求める。誘因には東北地方太平洋沖地震の際の計測震度を用いた。素因には、標高、傾斜方向、傾斜角、起伏量、微地形区分、河川からの距離を用いた。液状化発生を表すパス図の基本モデルの構造は、既往研究⁹⁾を参考に仮定した(図-2)。メッシュごとの液状化発生確率 S_q をベイジアンモデルにより求める¹⁰⁾。

ベイズの定理に従うと、液状化発生確率 S_q は以下のように記述できる。

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

ここで、式(1)を計算するためには、

$$P\{c_{ij}\} = \frac{N_{ij}}{A} \quad (2)$$

$$P\{c_{1j}, \dots, c_{mj}\} = \frac{N_x}{A} \quad (3)$$

$$P\{T_q\} = \frac{N_0}{A} \quad (4)$$

$$P\{T_q|c_{ij}\} = \frac{N_{d_{ij}}}{N_{ij}} \quad (5)$$

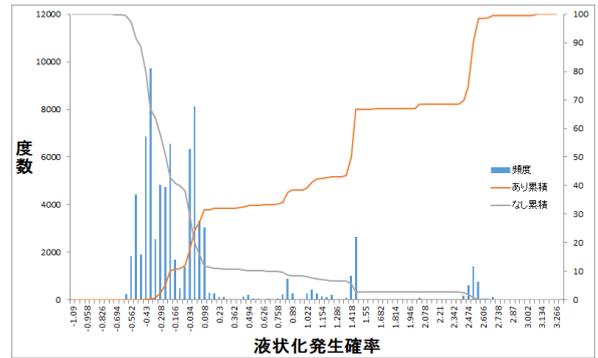
をそれぞれ代入する必要がある。ここで、 N_{ij} は*i*番目の素因におけるカテゴリ*j*のメッシュ数、 A は全メッシュ数、 N_x はメッシュ*q*における*i*~*m*番目の素因に対する全カテゴリにおいて等しい属性のメッシュ数、 N_0 は液状化発生のメッシュ数、 $N_{d_{ij}}$ は*i*番目の素因におけるカテゴリ*j*の領域に属する液状化発生のメッシュ数である。

共分散構造分析に基づき、液状化の発生を相関の高い要因を評価する。なお、ここでは液状化を引き起こす因果の強さを評価することを目的とし、標準化解を算出するために、観測変数は平均0、標準偏差1に標準化した。まず、図-2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1~2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。それぞれのモデルの液状化発生の的中率を算出し、最もの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液状化発生確率を、縦軸にその度数および累積頻度を示す¹⁰⁾。さらに、負極側から液状化発生地点の累積頻度曲線を、正極側から非液状化地点の累積頻度曲線を描く(図-3)。つまり、液状化発生地点をAグループ、非液状化発生地点をBグループとすると、

$$F_A(X) = 1 - \int_{-\infty}^X f_A(x) dx \quad \int_{-\infty}^{+\infty} f_A(x) dx = 1 \quad (6)$$

$$F_B(X) = \int_{-\infty}^X f_B(x) dx \quad \int_{-\infty}^{+\infty} f_B(x) dx = 1 \quad (7)$$

となり、累積度数分布 $F_A(X)$ 、 $F_B(X)$ の交点が判別の閾値 a となる。すなわち、液状化発生の危険がある領域は $X \geq a$ 、



モデル 1, 閾値: 0.045, 的中率: 79.73%

図-3 基本モデル基本モデルの累積頻度分布と的中率

表-1 精度評価のための混同行列

		実際のデータ	
		液状化あり	液状化なし
分析結果	液状化あり	TP (True Positive)	FP (False Positive)
	液状化なし	FN (False Negative)	TN (True Negative)

液状化発生の危険がない領域は $X < a$ となり、的中率は式(8)のようになる¹⁰⁾。

$$P = 1 - F_i(a) \quad (i = A, B) \quad (8)$$

なお、前述のように標準化解を算出することを目的とし、図-3の横軸(液状化発生確率 S_q :式(1))は標準化している。この結果、基本モデルの的中率が最も高い中率は79.7%となった。さらに、基本モデル(モデル1)のパスモデルによると、松岡ら⁸⁾の説明変数のうち傾斜角、微地形区分と液状化発生確率の相関係数が高いことがわかった。

4. 機械学習に基づく液状化発生地点の評価

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。機械学習に用いた変数は共分散構造分析の結果より傾斜角、微地形区分、計測震度を利用して分析を実行する。なお、カテゴリ変数である微地形区分をダミー変数化¹¹⁾して使用する。また、本研究では予測結果の評価方法として正答率、適合率、再現率、F値の4つの評価指標¹²⁾を導入する。各指標式は式(9)~式(12)の通りであり、式中のTP, FP, FN, TNは表-1のように定義される。

$$\text{正答率(Accuracy)} = \frac{TP+TN}{TP+FP+FN+TN} \quad (9)$$

$$\text{適合率(Precision)} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{再現率(Recall)} = \frac{TP}{TP+FN} \quad (11)$$

$$\text{F値(F-measure)} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

(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³⁾。

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、 \mathbf{w} は分離超平面の法線ベクトル、 ϕ は入力ベクトル \mathbf{x} を特徴空間 F へ非線形写像する関数 (式(14))、 b はスカラー変数を表す。

$$\min_{\mathbf{w}, b, \zeta} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \zeta_i$$

$$\text{s.t. } y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \zeta_i, i \in [n], \zeta_i \geq 0, i \in [n]$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (14)$$

さらに、本研究では非線形写像のための関数に、次式のRBFカーネルを用いる。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (15)$$

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³⁾。

本研究ではグリッドサーチ¹³⁾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 γ 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70%を学習データ、 γ を $10^{-5} \sim 10^5$ 、 C を $10^{-2} \sim 10^2$ の範囲に設定して、交差検証法($k=8$)によって最適なパラメータを求めた。また、危険箇所の見落としを減らすために液状化発生地点に重み付けを行った。重みの割合は液状化ありの地域となしの地域との割合より、21倍とした。

図-4に分析結果と実際の液状化発生状況との比較を行い、作成した液状化発生評価図と評価指標を示す。結果としては実際の液状化発生地点よりも過度に液状化発生の危険があるものと推定された。また、液状化地点の見落としが船橋市の沿岸部や佐倉市周辺で若干数発生した。しかし、重み付けをしたことによって、正答率と再現率の高い手法を構築することができた。

(2) ランダムフォレスト

ランダムフォレスト(RF)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、データの種別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⁴⁾。

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j) \quad (16)$$

ここで、 D_p 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 は個々の木を分割するノード数である。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¹⁴⁾を表している。

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

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。

図-4に液状化発生評価図および評価指標を示す。SVMに基づく結果と比較すると、RFによる再現率はSVMと比べて約5%低下したが、正答率、適合率、F値は、RFの分析結果の方が良い結果となった。

5. 液状化発生予測手法の適用性の検討

本研究で構築した機械学習による液状化発生地点の予測モデルを、東北地方太平洋沖地震の際の計測震度が約4.6~5.9で千葉県と同程度であった茨城県南部に適用し、モデルの汎用性の評価を行った。千葉県の際と同様に、茨城県の地形情報を国土数値情報⁵⁾と微地形区分⁶⁾より整理した。液状化発生地点は若松らがまとめた2011年の東北地方太平洋沖地震の際のデータを用いたり、分析に用いたメッシュ数の総数は38344メッシュであり、そのうち液状化発生地点は2568メッシュ存在した。結果を図-5に示す。RFを利用した予測モデルは、再現率の若干の低下が見受けられるものの、その他の評価指標はSVMを用いた予測モデルと比較して同程度であった。

次に別の地震を対象として、モデルの適用性に関する検討を行う。本研究で構築したSVMによる液状化発生予測モデルを、2016年に発生した熊本地震に適用する。県内112335メッシュを対象とし、そのうち液状化発生地点は1872メッシュであった。入力データは国土数値情報⁵⁾、微地形区分⁶⁾より整理し、熊本地震の計測震度は松岡¹⁵⁾による推定結果を用いた。液状化発生地点は若松らがまとめた2016年の熊本地震の際のデータを用いた¹⁶⁾。本研究で構築した機械学習に基づく液状化発生モデルの適用結果を図-6に示す。正

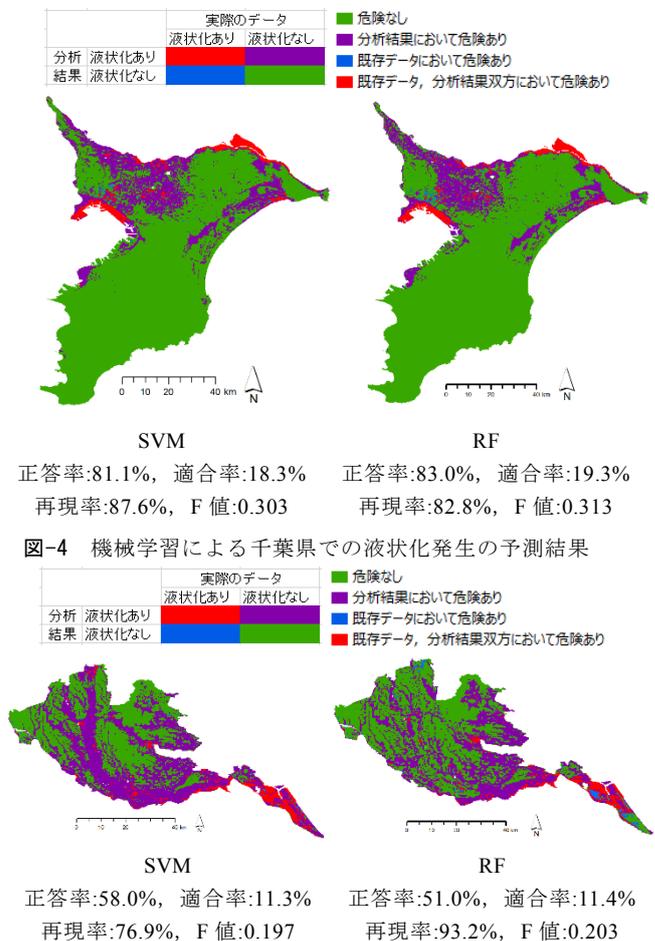
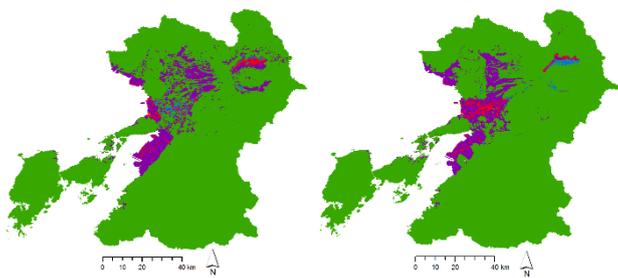


図-5 機械学習による茨城県南部での液状化発生の予測結果

	実際のデータ		■ 危険なし
	液状化あり	液状化なし	
分析結果	液状化あり	液状化なし	■ 分析結果において危険あり
結果	液状化あり	液状化なし	■ 既存データにおいて危険あり
			■ 既存データ、分析結果双方において危険あり



SVM 正答率:90.3%, 適合率:10.2% 再現率:61.3%, F 値:0.174
 RF 正答率:91.4%, 適合率:12.0% 再現率:66.2%, F 値:0.203

図-6 機械学習による熊本県での液状化発生の予測結果

答率は SVM, RF とともに 90%を超える値を算出することができたが、再現率が 60%台と低い値であった。

見落としが多く発生してしまった理由を考察するため、熊本県で液状化が発生したメッシュと、SVM, RF で液状化発生を見落としたメッシュの微地形区分を比較した。その結果、火山地や火山山麓地、火山性丘陵、砂礫質台地、扇状地で全ての液状化メッシュを見落とししていた。これは火山地や火山山麓地、火山性丘陵は千葉県に存在せず、砂礫質台地では液状化メッシュが全メッシュに対して極めて少数であり、扇状地では液状化が未発生であるためと考えられる。これらの微地形区分の地域を除いて再現率を算出し直すと、87.9%となった。このことから、学習データがすべての微地形区分を網羅するように整理されれば、SVM による液状化予測モデルは他の地震にも適用できる可能性があり、今後の検討課題である。RF での分析では砂礫質台地の液状化発生は予測することはできなかったものの、千葉県では液状化が発生していない火山山麓地や火山性丘陵、扇状地での液状化発生地点の一部を正しく分類することができた。しかし、後背湿地での液状化発生地点の見落としが非常に多く、再現率の低下の一因となっている。このことから RF による液状化予測モデルは微地形区分による影響よりも、計測震度や傾斜角の影響が大きいものと推察される。

6. まとめ

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を用いて、東北地方太平洋沖地震の際の千葉県を対象として液状化発生地点を予測する手法を構築した。共分散構造分析の結果によると、液状化の発生には微地形と傾斜角がとくに大きく影響することがわかった。SVM と RF で構築した予測モデルを比較すると、再現率は約 5%低下したが、正答率、適合率、F 値は、RF の分析結果の方が良い結果となった。

液状化予測モデルの適用性を評価するために他地域や他地震に適用した、千葉県と同程度の強さの地震動に曝された東北地方太平洋沖地震の際の茨城県南部では、SVM と RF での予測結果は同程度であった。別の地震への適用性の検討として、2016年熊本地震を対象とした分析を行った。その結果、SVMでの液状化予測モデルでは学習データで液状化が発生していなかった火山山麓地や扇状地での液状化

発生を正しく推定することができず、液状化地点の見落としがやや多くなった。一方、これらの微地形区分以外の液状化発生地点に関する再現率は高かった。RFでの液状化予測モデルは火山山麓地や扇状地での液状化発生の一部は正しく推定することができたものの、後背湿地での液状化発生地点の多くを推定することができなかった。

今後は、機械学習に用いる学習用データの見直しを行い、多くの微地形区分を網羅できるように再整理するなどして、液状化予測モデルの適用性の向上を図る必要がある。また、SVMとRFの長所を取り入れた二段階の機械学習を行ってみるなどして、さらなる液状化予測モデルの適用性の向上を図る必要がある。さらに、液状化履歴の有無や地盤改良等の液状化対策が行われているか否かを考慮できるようにパラメータの見直しも行っていきたい。

参考文献

- 1) 若松加寿江, 先名重樹: 2011年東北地方太平洋沖地震による関東地方の液状化発生と土地条件, 日本地震工学会論文集, Vol.15, No.2, pp.25-44, 2015.
- 2) 豊田秀樹: 共分散構造分析[R編], 東京図書, 2014.
- 3) 竹内一郎, 鳥山昌幸: サポートベクターマシン (機械学習プロフェッショナルシリーズ), 講談社, 2015.
- 4) 波部齊: 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.
- 5) 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/>
- 6) 若松加寿江, 松岡昌志: 全国統一基準による地形・地盤分類 250m メッシュマップの構築とその利用, 日本地震工学会誌, No.18, pp.35-38, 2013.
- 7) 産業技術総合研究所: 地震動マップ即時推定システム, <https://gbank.gsj.jp/Quake/QuakeMap/index.html>
- 8) 松岡昌志, 若松加寿江, 橋本光史: 地形・地盤分類 250m メッシュマップに基づく液状化危険度の推定手法, 日本地震工学会論文集, Vol. 11, No. 2, pp.20-39, 2011.
- 9) 小島尚人, 大林成行, 青木太: 共分散構造分析を導入した斜面崩壊危険箇所評価アルゴリズムの構築, 土木学会論文集, No.714/VI-56, pp.79-93, 2002.
- 10) 大林成行, 小島尚人, Chang-Jo F.Chung: 斜面安定性評価モデルの精度比較とその実用化への提案, 土木学会論文集, No.630/VI-44, pp.77-89, 1999.
- 11) 豊澤栄治: 最終回/ダミー変数を使って上手にデータをまとめ、グループごとの特徴を把握しよう【R入門講座】, <http://markezine.jp/article/detail/20790>
- 12) 平井 有三: はじめてのパターン認識, 森北出版, 2012.
- 13) 荒川正幹, 宮尾知幸, 船津公人: ドラッグライクネスモデルの構築とその可視化, Journal of Computer Aided Chemistry, Vol. 9, pp. 70-80, 2008.
- 14) 山岡 啓介: ランダムフォレスト, 映像情報メディア学会誌, Vol. 66, No. 7, pp. 573-575, 2012
- 15) 松岡昌志: 強震観測記録に基づく 2016年熊本地震の地震動分布の推定, 地域安全学会梗概集, No.42, pp.23-26, 2018.
- 16) 若松加寿江, 先名重樹, 小澤京子: 平成 28 年 (2016 年) 熊本地震による液状化発生の特性, 日本地震工学会論文集, Vol. 17, No. 4, pp.81-100, 2017