

共分散構造分析とサポートベクターマシンを用いた液状化発生地点の特徴の評価

柳瀬 匡雄¹・丸山 喜久²

¹学生会員 千葉大学大学院融合理工学府 博士前期課程（〒263-8522 千葉県千葉市稲毛区弥生町1-33）

E-mail: 14tm0311@chiba-u.jp

²正会員 千葉大学大学院工学研究院 准教授（〒263-8522 千葉県千葉市稲毛区弥生町1-33）

E-mail: ymaruyam@tu.chiba-u.ac.jp

本研究では若松・先名（2015）がとりまとめた東北地方太平洋沖地震における液状化発生地点データを利用し、千葉県内の液状化発生地点と地震動強さ、地形的特徴の因果関係を共分散構造分析とサポートベクターマシン（SVM）による分析を用いて評価した。さらに、千葉県内のデータを用いて構築したSVMによる液状化発生の判別モデルを、茨城県南部を対象に適用性を検討した。共分散構造分析での結果は、液状化発生の危険性が高い地点が本来の液状化発生地点に比べて過剰に抽出された。一方、SVMによる液状化発生地点の判別精度は非常に高く、液状化の危険度評価に有用と考えられる。

Key Words : soil liquefaction, covariance structure analysis, support vector machine, 2011 Tohoku earthquake

1. はじめに

2011年東北地方太平洋沖地震によって、青森県から神奈川県までのおよそ南北650kmの広域な範囲で液状化が発生した¹⁾。液状化は、土構造物、橋梁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。また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²⁾、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有意義と考えられる。

関東地方は大河川の沖積作用による低地や埋立地の面積が多いため、東北地方太平洋沖地震では液状化が多数発生した。一方、東北地方は津波の襲来によって液状化の痕跡が消えてしまったこともあるが、海岸近くまで山地・丘陵が迫っているため、液状化による被害が少なかった。このように液状化の発生には、地形的特徴の違いが大きく影響したものと考えられる³⁾。

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)⁴⁾と、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)⁴⁾を用いて、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。

本研究では、千葉県を対象に分析を行った。使用したデータ（表-1）は国土交通省が整備する国土数値情報⁵⁾、微地形区分⁶⁾である。前者には標高、傾斜角などの地形データが5次メッシュ（250mメッシュ）毎に整備されている。後者には微地形が5次メッシュ（250mメッシュ）毎に格納されている。本研究で用いる地形的特徴データは、標高（図-1）、傾斜方向、傾斜角、起伏量、微地形、河川からの距離であり、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。地形的特徴として傾斜を考慮したり、地震動指標に計測震度を用いたことは、松岡ら⁸⁾による液状化危険度の推定手法を参考にした。また、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を用いた。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-2）。

2. 共分散構造分析に基づく液状化発生確率の評価

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する。さらに、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を求める。まず、誘因には東北地方

表-1 カテゴリごとのメッシュ数

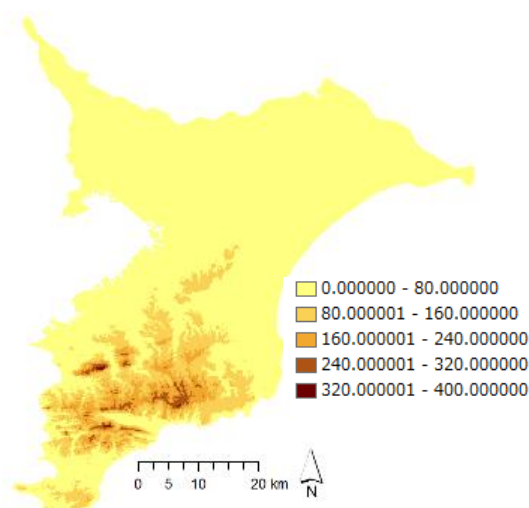
[illegible]

図-1 千葉県の標高分布

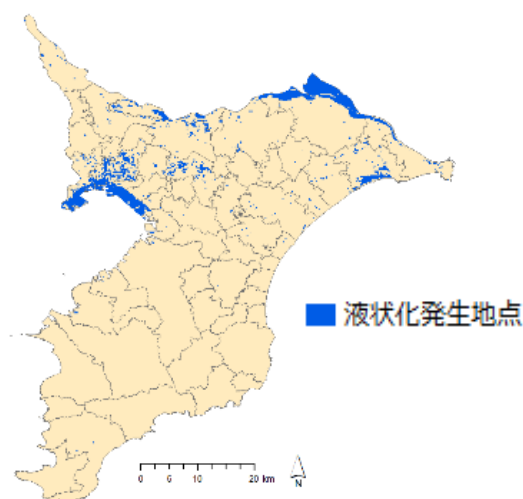


図-2 千葉県の液状化発生地点¹⁾

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（メッシュ内の標高差の最大値）、微地形区分、河川からの距離を用いた。共分散構造分析では、直接観測できる変数（観測変数）に加えて、内容の似通った観測変数をまとめて構成概念を表現する潜在変数を扱うことができる。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⁾を参考にした（図-3）。共分散構造分析にあたり、メッシュごとの液状化発生確率 S_q をベイズアンモデルにより求める¹⁰⁾。メッシュ q において液状化が発生すると想定される事象を $\{T_q\}$ とし、 $P\{T_q\}$ はその確率を表す。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_{ij} ($i = 1, \dots, m$, $j = 1, \dots, n$)と記述する。

ベイズの定理に従い、液状化発生確率 S_q は以下のよう
に記述できる.

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

ここで、各素因が独立事象であると仮定すると、

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

となり、式(2)を式(1)に代入すると、

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

となる. ここで, 式(3)を計算するためには,

$$P\{c_{ij}\} = \frac{N_{ij}}{A} \quad (4)$$

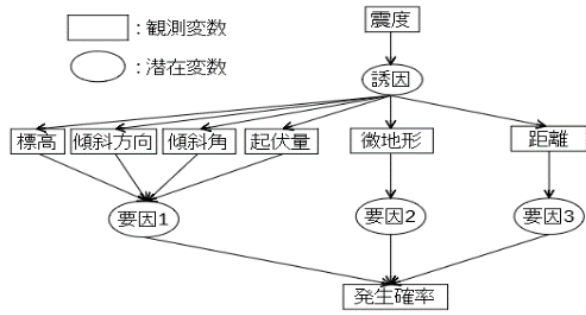


図-3 本研究で仮定した液状化発生の基本モデル

表-2 パスモデルの適合度指標

モデル名	削除した要因	RMSEA	CFI	TLI
モデル1	-	0.15	0.868	0.754
モデル2	標高	0.149	0.869	0.725
モデル3	傾斜方向	0.127	0.912	0.815
モデル4	傾斜角	0.194	0.781	0.54
モデル5	起伏量	0.152	0.875	0.737
モデル6	微地形	0.219	0.795	0.57
モデル7	距離	0.191	0.853	0.692
モデル8	震度	0.095	0.968	0.916
モデル9	標高+傾斜方向	0.3	0.746	0.365
モデル10	標高+傾斜角	0.211	0.705	0.263
モデル11	標高+起伏量	0.169	0.864	0.659
モデル12	標高+微地形	0.264	0.726	0.316
モデル13	標高+距離	0.201	0.85	0.624
モデル14	標高+震度	0.021	0.999	0.995
モデル15	傾斜方向+傾斜角	0.426	0.573	-0.067
モデル16	傾斜方向+起伏量	0.305	0.752	0.38
モデル17	傾斜方向+微地形	0.339	0.771	0.426
モデル18	傾斜方向+距離	0.361	0.749	0.372
モデル19	傾斜方向+震度	0.092	0.969	0.928
モデル20	傾斜角+起伏量	0.213	0.739	0.348
モデル21	傾斜角+微地形	0.363	0.603	0.008
モデル22	傾斜角+距離	0.239	0.754	0.53
モデル23	傾斜角+震度	0.037	0.996	0.985
モデル24	起伏量+微地形	0.259	0.751	0.378
モデル25	起伏量+距離	0.206	0.855	0.639
モデル26	起伏量+震度	0.026	0.998	0.993
モデル27	微地形+距離	0.309	0.763	0.492
モデル28	微地形+震度	0.04	0.997	0.989
モデル29	距離+震度	0.131	0.97	0.886

は液状化発生のメッシュ数、 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¹¹⁾。まず、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1～2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にまとめる。

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以上の値を取り、値が小さいほどモデルの適合が良いとされる。一般にRMSEAは0に近いほど適合が良いモデルとされ、0.08以下でモデルが採択される¹²⁾。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。独立モデルとは、あるデータの分析をする場合に、想定し得る中で最も適合の悪いモデルのことである。これらは、1に近いほど適合が良いと判断される¹³⁾。

表-2の結果から、基本モデル(モデル1)の適合度は比較的良好であるが、誘因である計測震度を除いたモデル8、14、26の適合度が高い。これは、千葉県内で液状化が発生した地点の約80%が震度5強であり、震度階級に大きな差がなかったことが影響していると考えられる。また、計測震度を残したモデルの中では、モデル4、13の適合度が高かった。したがって、基本モデルに加えてこれらの5つのモデルについて詳細な検討を行う。

図-4に基本モデルとこれらの5つのモデルのパス図とパス係数(相関係数)を示す。それぞれのパスモデルの相関係数から、液状化発生には要因2(微地形区分)が最も大きく影響しており、次いで要因1の影響が大きかった。また、要因3(河川からの距離)の影響は小さい。また、要因1の中では傾斜角の相関係数が最も大きかった。

図-4の結果をもとに液状化発生地点を推定し、その的中率を評価する。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-5)。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、

$$P\{c_{1j}, \dots, c_{mj}\} = \frac{N_X}{A} \quad (5)$$

$$P\{T_q\} = \frac{N_0}{A} \quad (6)$$

$$P\{T_q|c_{ij}\} = \frac{N_d}{N_{ij}} \quad (7)$$

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*～*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

$$F_A(X) = 1 - \int_{-\infty}^X f_A(x)dx \quad (\int_{-\infty}^{+\infty} f_A(x)dx = 1) \quad (8)$$

$$F_B(X) = \int_{-\infty}^X f_B(x)dx \quad (\int_{-\infty}^{+\infty} f_B(x)dx = 1) \quad (9)$$

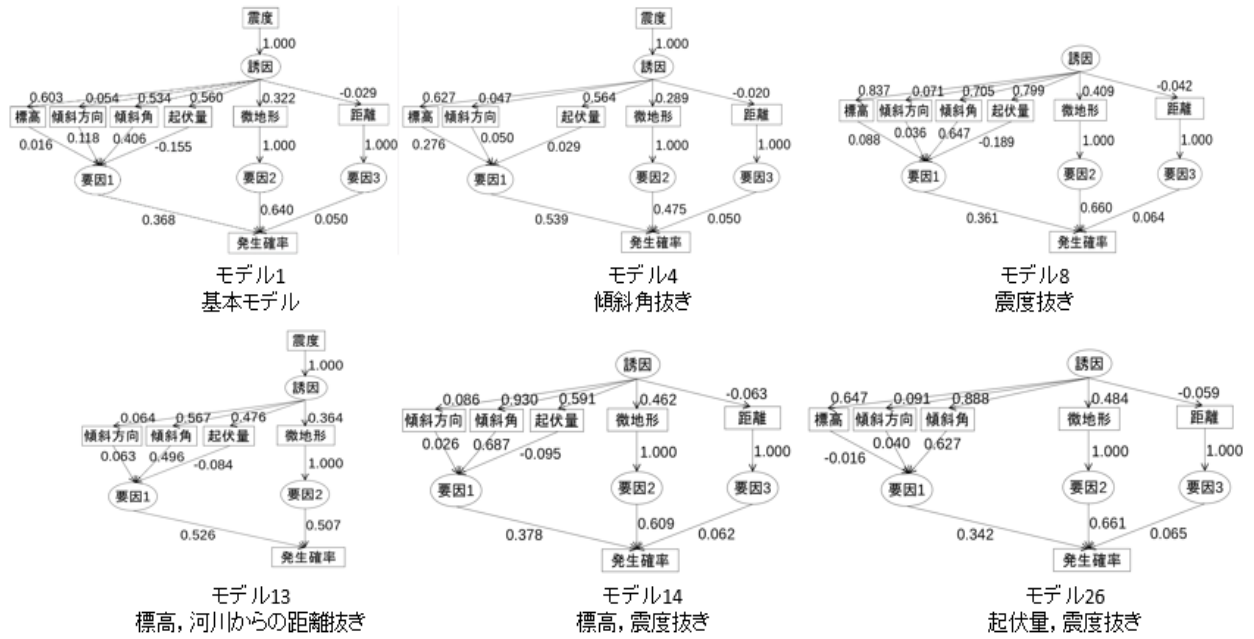


図-4 採用した6ケースのパスモデル

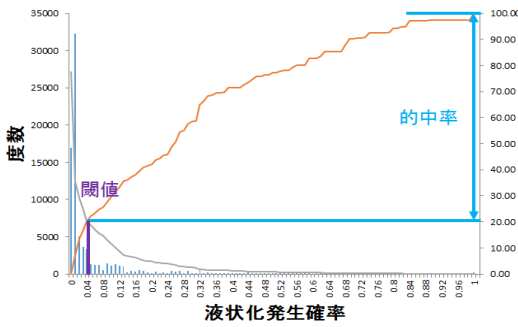


図-5 液状化発生確率の累積頻度分布と的中率

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⁷⁾。すなわち、液状化発生の危険がある領域は $X \geq a$ 、液状化発生の危険がない領域は $X < a$ となり、的中率は式(10)のようになる¹⁰⁾。

$$P = 1 - F_i(a) \quad (i=A,B) \quad (10)$$

各モデルの的中率を図-6に示す。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-6の横軸（液状化発生確率 S_q ：式(1)）は標準化している。基本モデルの的中率が最も高く、79.7%となった。

3. サポートベクターマシンによる2クラス分類

共分散構造分析の結果を踏まえて、本研究ではサポートベクターマシン（SVM）を用いた分析を実施した。SVMとは、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器を構成するための学習方法とされ、認識性能の優れた学習モデルの一つといわれている⁴⁾。本研究では、Gaussianカーネルを用いたSVMを行う。さらに、共分散構造分析の結果より、基本モデル（モデル1）のパスモデルで液状化発生確率に大きく影響を与えている地形的要因は、傾斜角と微地形区分であることがわかった。したがって、SVMにおいては傾斜角、微地形区分、計測震度を利用して分析を実行する。なお、SVMではカテゴリ変数である微地形区分をダミー変数化¹⁴⁾して使用する。

SVMによる2クラス分類では、式(11)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(12)）、 b はスカラー変数を表す⁴⁾。

$$\begin{aligned} \min_{\mathbf{w}, b, \zeta} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \zeta_i \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \zeta_i, i \in [n], \zeta_i \geq 0, i \in [n] \end{aligned} \quad (11)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (12)$$

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。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (13)$$

式(13)の γ と式(11)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⁴⁾。

最も分類精度が高くなるハイパーパラメータを求める

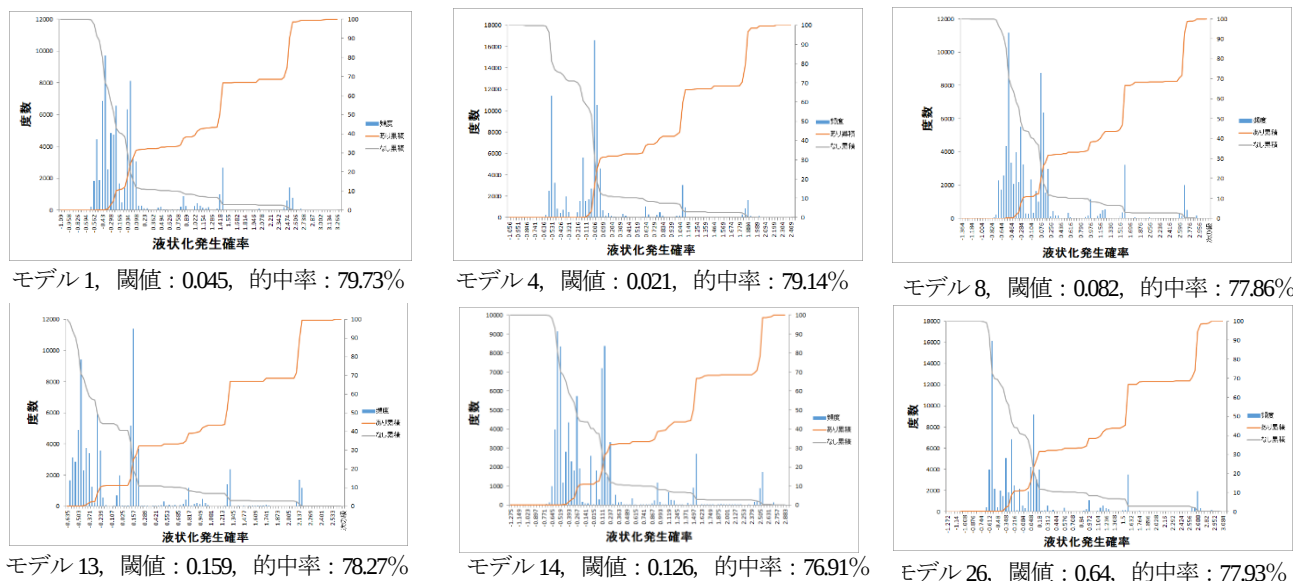


図-6 採用した6ケースの液状化抽出の的中率

表-3 10%学習(重み付きなし)での分析結果と実際の液状化発生地点の比較

		実際のデータ	
		危険あり	危険なし
分析結果	危険あり	57	3
	危険なし	3645	75990

ために、グリッドサーチ¹⁹⁾という手法を用いる。グリッドサーチとは、 γ とCの値がとる範囲を任意に設定して、交差検証法により最適な γ とCの値を求める手法である。まず本研究では、全データの10%を学習データ、 γ を 10^{-5} ~ 10^5 の範囲、Cを 10^{-2} ~ 10^2 の範囲に設定して、10分割の交差検証法により、適切なパラメータを求めた。この結果、 $\gamma=2.51 \times 10^2$ 、 $C=1.0$ が最適解となった。

求めたハイパーパラメータを用いて分類器を作成し、液状化地点との比較を行ったものを表-3に示す。表-3では、実際の液状化発生地点よりも、SVMによる分類結果が過小評価となった。これには、学習データ中の液状化発生地点と非液状化地点数の不均衡が影響しているものと考えられる¹⁹⁾。そこで、液状化発生危険地点のメッシュに重み付けをすることにより、液状化地点を多く抽出できるようにする。全データの10%を学習データとし、液状化発生地点と非発生地点のデータ数の比に基づいて、液状化発生地点に21倍の重み付けしたところ、最適なパラメータは $\gamma=1.25 \times 10^3$ 、 $C=0.1$ となった。SVMによる分類結果と実際のデータとの比較結果を表-4に示す。液状化発生地点の見落としが大きく改善され、的中率も

表-4 10%学習(重み付きあり)での分析結果と実際の液状化発生地点の比較

		実際のデータ	
		危険あり	危険なし
分析結果	危険あり	3695	0
	危険なし	7	75993

非常に高くなったことがわかる。

4. 液状化発生評価結果の比較

実際の液状化発生地点のデータと共分散構造分析およびSVMにより算出した液状化発生評価結果を比較し、結果をGIS上に表示した(図-7、図-8)。共分散構造分析(基本モデル)による液状化地点の的中精度(液状化発生地点と非発生地点の的中率)は80.5%となった。千葉県の北部を中心に、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。具体的には、県北部のROOM台地の標高が低い地点で、液状化の発生の危険性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数多く発生してしまった。

次に、SVMによる液状化発生評価結果(図-8)では、重み付けをしたことによって、液状化地点の見落としが激減した。さらに、的中率も99.9%と非常に良好な結果が得られた。

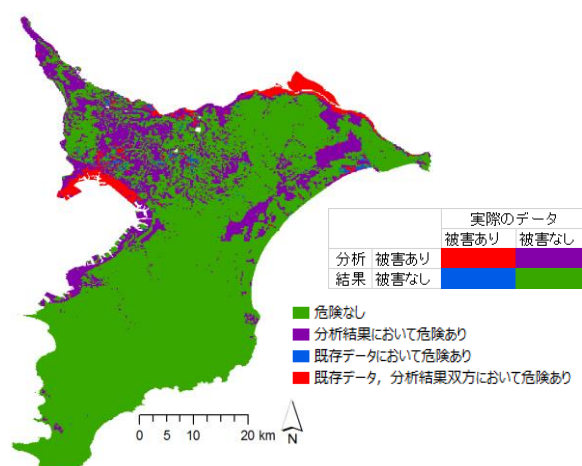


図-7 共分散構造分析による液状化発生評価結果

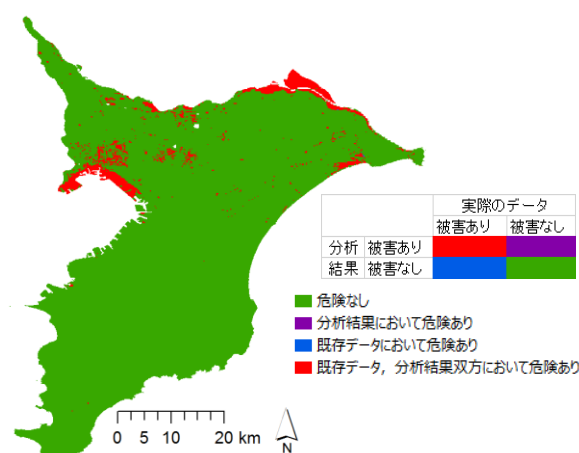


図-8 SVMによる液状化発生評価結果

表-5 茨城県南部でのSVMによる分析結果と実際の液状化発生地点の比較

		実際のデータ	
		危険あり	危険なし
分 結 析 果	危険あり	2536	0
	危険なし	32	35776

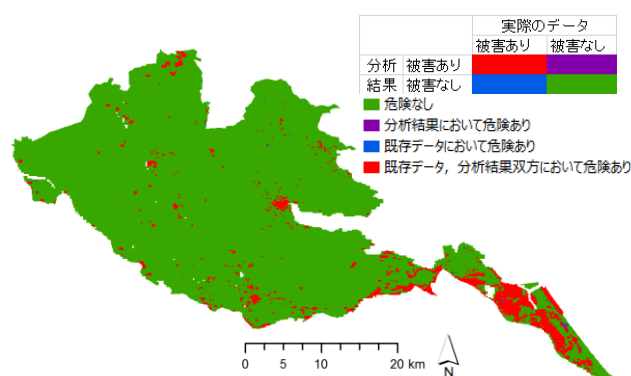


図-9 茨城県南部での SVM による液状化発生評価結果

本研究で構築したSVMによる液状化発生地点の分類器を、東北地方太平洋沖地震における計測震度が千葉県と同程度であった茨城県南部に適用し、液状化発生評価を行った。千葉県の際と同様に、茨城県の地形情報を国土数値情報⁹⁾と若松・松岡の微地形区分⁹⁾より取得し、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¹⁾。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。千葉県のデータを用いて構築したSVM（液状化発生地点に21倍の重み付けあり）による液状化分類器の適用結果を表-5、図-9に示す。

表-5、図-9より、千葉県で構築したSVM分類器であるため、液状化発生地点の見落としが増加してしましたが、茨城県南部に適用した場合の精度は99%と非常に高く、計測震度が同程度であれば他地域への適用も期待できる。

5. 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ことを試みた。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。構築されたパスモデルによる分類結果は、実際に液状化が発生していない多くの地点で過剰に液状化の危険性があるように推定された。一方、SVMによる液状化発生の評価結果は、非常に高い的中率を得ることができた。さらに、千葉県のデータで構築したSVMによる分類器を茨城県南部について適用したところ、良好な分類結果が得られた。

以上の結果から、計測震度が同程度の地域であれば、本研究の手法で高い精度で液状化発生地点を予測できる可能性がある。今後は、計測震度がより大きな地域を対象とした再検討、他の地震への適用性の検討などが必要である。

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) 内閣府 (防災担当): これまでの首都直下地震対策について, http://www.bousai.go.jp/jishin/syuto/taisaku_wg/1/pdf/1.pdf

- 3) 豊田秀樹：共分散構造分析[R編]，東京図書，2014.
- 4) 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.
- 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) 豊田秀樹：共分散構造分析 [疑問編] ー構造方程式モデリングー，朝倉書店，2003.
- 12) 岡浩一郎，石井香織，柴田愛：日本人成人の身体活動に影響を及ぼす心理的，社会的，環境的要因の共分散構造分析，体力科学，Vol.60, No. 1，pp.89-97, 2011.
- 13) 廣野元久：グラフィカルモデリングのための G-GM & L-GM データ解析システム，計算機統計学，Vol. 15, No. 1, pp. 63-74, 2003.
- 14) 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>
- 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の構築とその可視化，Journal of Computer Aided Chemistry，Vol. 9, pp. 70-80, 2008.
- 16) 齋竹良介，荒井幸代：実被害情報を活用した水道管被害箇所のオンライン再推定，人工知能学会全国大会論文集，Paper No. 1M3-OS-24b-4, 2015.

EVALUATION OF THE TOPOGRAPHIC CHARACTERISTICS AT THE LIQUEFIED AREAS BASED ON COVARIANCE STRUCTURE ANALYSIS AND SUPPORT VECTOR MACHINE

Masao YANASE and Yoshihisa MARUYAMA

This study tries to develop a method to estimate the locations of soil liquefaction occurrences based on the covariance structure analysis and support vector machine. 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, a dataset of liquefaction occurrences after the 2011 off the Pacific Coast of Tohoku Earthquake compiled by Wakamatsu and Senna (2015) was employed in this study. This study evaluated the topographic characteristics at the sites of liquefaction in Chiba Prefecture and the southern part of Ibaraki Prefecture.