

地震動の継続時間を考慮した機械学習に基づく 液状化発生予測に関する基礎的検討

中山 洋斗¹・劉 ウェン²・丸山 喜久²

¹学生会員 千葉大学博士前期課程 大学院融合理工学府 (〒263-8522 千葉県千葉市稲毛区弥生町 1-33)
E-mail: afta6895@chiba-u.jp

²正会員 千葉大学 大学院工学研究院 (〒263-8522 千葉県千葉市稲毛区弥生町 1-33)
E-mail: wen.liu@chiba-u.jp, ymaruyam@faculty.chiba-u.jp

東北地方太平洋沖地震では、東日本の広域において液状化が発生した。現在の液状化の予測手法では、地盤調査データを利用したものが高精度であり一般的である。しかし、広域にこの高精度な評価を適用することは困難である。そのため、道路や水道、都市ガスのようなライフラインネットワークの被害予測には、簡便に面的な評価を行える手法の確立が望まれる。そこで、本研究では機械学習のうちサポートベクターマシンとランダムフォレストを用いて、液状化発生予測モデルを構築することを目的とする。既往研究の検討結果を踏まえて、液状化の発生に影響がある地震動継続時間を変数として新たに加えた。さらに、液状化の発生傾向によってモデルを微地形区分に基づきグループ分けし、予測精度の高いモデルの作成を目指した。

Key Words: Strong motion duration, liquefaction, support vector machine, random forests

1. はじめに

2011年の東北地方太平洋沖地震では、青森県から神奈川県までの広範囲に及び液状化が発生した¹⁾。液状化が発生すると、一般の構造物のほかにも橋梁やライフライン施設にも甚大な被害を与える。都市ガスのリアルタイム地震防災システムに液状化予測機能が搭載されている²⁾ことを踏まえても、液状化が発生した地点を迅速に予測することは防災直後の緊急対応に大きく貢献できるものと期待される。地点ごとの液状化の危険度評価は、地盤調査データから液状化指数 (P_L 値)³⁾によって実施されることが多く、高精度な評価が可能である。しかし、詳細な地盤調査データを広域に取得することは困難であることから、道路や水道、都市ガスのようなライフラインネットワークの被害予測には、簡便に面的な評価を行える手法の確立が望まれる。例えば松岡ら⁴⁾は、微地形区分⁵⁾ごとに液状化発生危険度を評価する方法を構築している。また、桑原・松岡⁶⁾は、機械学習を用いて日本全国の液状化危険度を評価する手法を構築した。これらの研究は、液状化ハザードマップの作成に有用である。

柳瀬・丸山⁷⁾は、若松・先名⁸⁾が取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、千葉県内の液状化地点と地形的特徴の因果関係を共分散

構造解析で評価し、計測震度と微地形区分のほかにも傾斜角も液状化の発生確率に影響があることを示した。さらに、この分析結果に基づき、サポートベクターマシン (SVM) による液状化発生の予測モデルを作成した。

柳瀬・丸山⁷⁾では、SVMに入力する地震動に関する変数として計測震度だけを用いている。一方、液状化発生率と地震動の継続時間には相関関係があると評価されており、これを用いた液状化発生率の予測法も提案されている⁹⁾。そこで本研究では、広域性と即時性を指向した液状化発生予測モデルの構築を目指し、地震動の継続時間を考慮した機械学習を実行する。これによって、既往の柳瀬・丸山⁷⁾よりも精度が高い予測モデルの構築を目標とする。

2. 本研究で用いたデータ

本研究では、柳瀬・丸山⁷⁾と同様に、東北地方太平洋沖地震の際の千葉県を対象とする。東北地方太平洋沖地震では前震、余震を含め、複数の地震が発生しているが、液状化の発生には本震の影響が最も大きかったものと仮定する。本研究で使用したデータは、防災科学技術研究所の地震ハザードステーション (J-SHIS)¹⁰⁾の微地形区



図-1 千葉県液状化地点⁹⁾

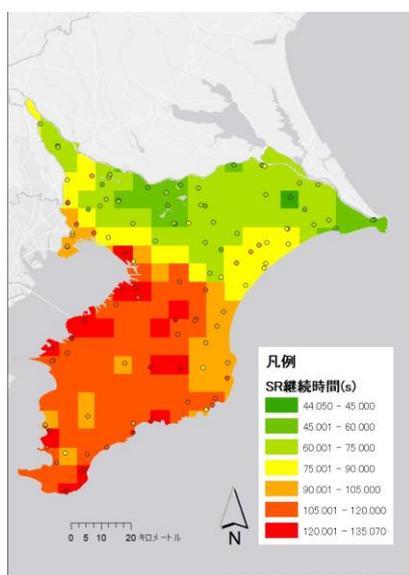


図-2 SR継続時間の分布

分、防災科学技術研究所の強震観測網 (K-NET, KiK-net)¹¹⁾の加速度記録、千葉県強震観測網¹²⁾の加速度記録、および産業総合技術研究所の QuiQuake¹³⁾による計測震度分布である。このうち微地形区分と計測震度分布は5次メッシュ(250mメッシュ)ごとに格納されている。また、液状化発生地点には、図-1に示す若松・先名⁹⁾がまとめた5次メッシュごとに集計されたデータを使用する。

加速度記録は地震動の継続時間を算出するために使用した。地震動の継続時間は、地動加速度の累積パワー曲線の占有時間から算出されるSR (Significant/Relative) 継続時間¹⁴⁾とし、累積パワーが5~95%の範囲の時間を地震動の継続時間とした。東北地方太平洋沖地震の本震の地震動の継続時間を防災科学技術研究所の強震観測点37地点と千葉県強震観測網の76地点について計算した。さらに、GIS上の内挿補間の手法の1つであるIDW法

表-1 精度評価のための混合行列

		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)

(Inverse Distance Weighted: 逆距離加重法) によって空間補間し、250mメッシュごとの地震動継続時間の推定値を算出した(図-2)。

3. 機械学習に基づく液状化発生予測モデル

本研究で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サポートベクターマシン (SVM)¹⁵⁾と、複数の決定木を用いて分類を行う機械学習アルゴリズムであるランダムフォレスト (RF)¹⁶⁾の2種類の機械学習を用いた。千葉県の79812メッシュを学習データとテストデータに2分割し、液状化発生予測モデルの構築と精度検証を行った。モデルの精度評価指標には、正解率 (Accuracy: *Acc*)、適合率 (Precision: *Pre*)、再現率 (Recall: *Rec*)、適合率と再現率の調和平均であるF値の4つの指標¹⁷⁾を用いる。各指標値は式(1)-(4)の通りであり、式中のTP, FP, FN, TNは表-1の通りである。

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Pre = \frac{TP}{TP+FP} \quad (2)$$

$$Rec = \frac{TP}{TP+FN} \quad (3)$$

$$F = \frac{2 \cdot Precision \cdot Recall}{precision+Recall} \quad (4)$$

正解率は液状化発生地点の有無を正しく分類した割合である。適合率は、液状化発生地点であると予測した地点のうち、実際に液状化が発生している地点の割合である。再現率は、実際に液状化が発生した地点のうち、液状化が発生したと予測されたものの割合である。この適合率と再現率はトレードオフの関係であり、どちらか一方が増加するともう一方は低下する。F値は適合率と再現率の調和平均であるため、2クラス分類においてよく用いられる指標である。

(1) サポートベクターマシン

サポートベクターマシン (SVM) では、ハイパーパ

表-2 SVM1 と SVM2 の混同行列と精度

SVM1		実際のデータ	
		液状化あり	液状化なし
被害予測	液状化あり	1917	533
	液状化なし	1439	67943
正答率	0.973	再現率	0.571
適合率	0.782	F値	0.660

SVM2		実際のデータ	
		液状化あり	液状化なし
被害予測	液状化あり	1369	505
	液状化なし	1987	67971
正答率	0.965	再現率	0.408
適合率	0.731	F値	0.524

表-3 液状化の発生傾向に基づく微地形のグループ

区分	微地形区分
グループ1	埋立地, 干拓地, 旧河道, 自然堤防, 河原
グループ2	扇状地, 砂州・砂礫洲
グループ3	後背湿地, 三角州・海岸低地, 砂丘, 砂州・砂丘間低地
グループ4	谷底低地, 砂礫台地
グループ5	山地, 山麓地, 丘陵, 火山地, 火山山麓地, 火山性丘陵, 岩石台地, ローム台地, 岩礁

ラメータである σ , C を決定する必要がある。 σ と C の値の範囲を任意に設定し, すべての組み合わせを試し, 交差検証法により最も分類精度が高くなる σ と C の組み合わせを求める。本研究では σ の範囲を $10^5 \sim 10^6$, C を $10^2 \sim 10^2$ に設定して, 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。

まず, 液状化発生予測モデルに地震動の継続時間を新たに導入したことの効果を評価するため, 地震動の継続時間を説明変数に含めたモデル (SVM1) と含めないモデル (SVM2) を作成した。学習データの割合は全体の 10% とし, 5 分割交差検証法により最も分類精度の高くなるハイパーパラメータの組み合わせを定めた。その結果, SVM1 のハイパーパラメータは $\sigma = 10^{11}$, $C = 10^{06}$, SVM2 のハイパーパラメータは $\sigma = 10^{21}$, $C = 10^{03}$ となった。なお, 学習データ中の液状化発生地点数 (3754) と非液状化地点数 (76058) が大きく異なるため, これらの比に基づき液状化発生地点に 20 倍の重みづけを行った。分類器の作成後, 学習に用いなかった残りの 90% のテストデータに対して SVM1 と SVM2 を適用し, 精度評価を行った。その結果を表-2 に示す。地震動の継続時間を加えることで液状化の見落としを大幅に減らすことができ, F 値が約 0.13 向上した。このことから, SVM に

表-4 SVM3 のハイパーパラメータと重み

	重み	σ	C
グループ1	3	$10^{2.2}$	$10^{1.9}$
グループ2	12	$10^{5.0}$	$10^{-0.4}$
グループ3	7	$10^{3.8}$	$10^{-0.2}$
グループ4	71	$10^{3.4}$	10^0
グループ5	70	$10^{2.6}$	10^0

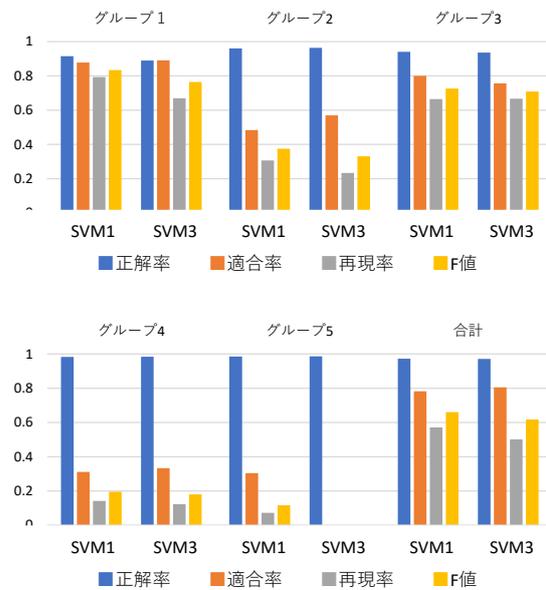


図-3 SVM1 と SVM3 の精度評価指標の比較

よる液状化発生予測モデルの説明変数に地震動の継続時間を含めることは有効であることがわかる。

さらに液状化発生予測モデルの精度を向上させるため, 既往研究²⁹⁾に基づき, 液状化の発生傾向別のグループ1~5ごとに液状化発生予測モデルを構築 (SVM3) し, グループ分けをしない場合の SVM1 と比較する。液状化の発生傾向別の微地形の区分は表-3の通りである。学習データの割合は, SVM1, SVM2 と同様に全体の 10% とする。ハイパーパラメータの決定方法は SVM1, SVM2 と同条件とし, データの不均衡性のための重みづけは, 各グループの液状化発生地点数と非液状化地点数の比とした。学習データの分類精度を最大とするハイパーパラメータと各グループの重みの値は表-4の通りである。

残りの 90% のテストデータに対して SVM3 を適用し, 精度評価を行った。図-3 に, SVM1 (グループ分けなし) と SVM3 (グループ分けあり) の精度指標値を比較する。液状化の発生が多いグループ1とグループ3では, SVM1 と SVM3 のどちらも精度が高いが, SVM1 の方がどの精度指標もわずかに大きな値を示した。なお, SVM3 のグループ5については, 液状化であると予測した地点がな

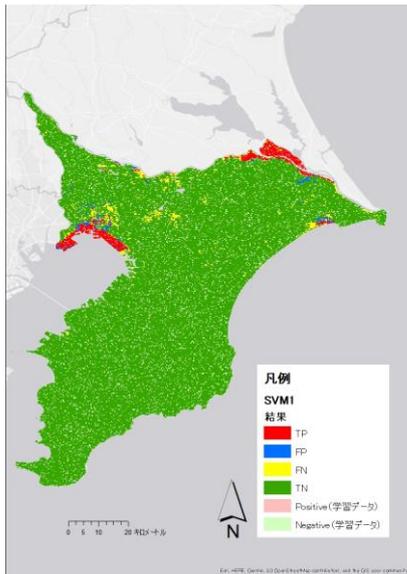


図4 SVM1の予測結果

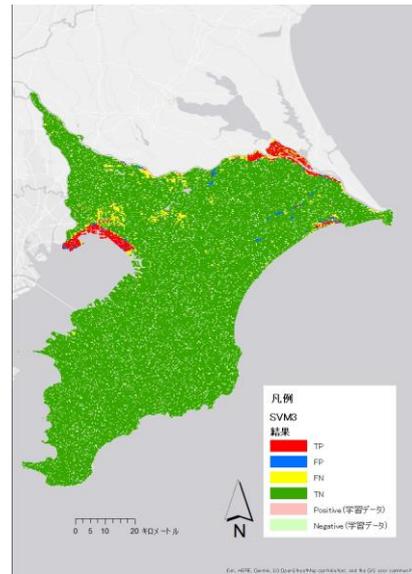


図5 SVM3の予測結果

かった。このため、再現率は0、適合率とF値は計算できない。グループ1~5の合計では、SVM3の方が適合率が高いが、液状化の見落としがSVM1と比べて多く、適合率とF値はSVM1よりも低下した。

全体の90%を占めるテストデータに対する分類結果を図示したものが図4、図5である。東京湾岸や利根川流域の液状化が集中している地域では液状化の予測が的中しているものの、内陸部の液状化が点在している地域では見落としが多くみられた。また、SVM1のほうが内陸部での液状化を比較的多く予測している。

(2) ランダムフォレスト

ランダムフォレスト (RF) による液状化発生予測モデルの構築を検討した。RFでは、決定木を成長させる際に使用する特徴量の数 $mtry$ を指定する必要がある。本研究では、この $mtry$ をハイパーパラメータとして予測モデルを構成する。前述のSVMと同様に5分割交差検証法によるグリッドリサーチにより、学習データに対する分類精度が最も高くなる $mtry$ を定める。また、RFではデータの不均衡性による重みづけは行わない。学習データとテストデータの割合は7:3とした。

まず、液状化発生予測モデルの説明変数として地震動の継続時間を導入したことの有効性を検討する。地震動の継続時間を説明変数として含めたRF1と含めないRF2を構築した。学習データに対する分類精度が最大となるハイパーパラメータは、RF1が $mtry = 24$ 、RF2が $mtry = 23$ となった。残りの30%のテストデータにRF1とRF2を適用し、精度評価を行った結果を表5に示す。RFでもSVMと同様に地震動の継続時間を説明変数に加えることで液状化の見落としを大幅に減らすことができた。

表5 RF1とRF2の混同行列

RF1		実際のデータ	
		液状化あり	液状化なし
被害予測	液状化あり	738	192
	液状化なし	399	22617

正答率	0.975	再現率	0.649
適合率	0.794	F値	0.714

RF2		実際のデータ	
		液状化あり	液状化なし
被害予測	液状化あり	476	133
	液状化なし	661	22676

正答率	0.967	再現率	0.782
適合率	0.419	F値	0.545

RF1はRF2よりも再現率は低下したものの、正解率、適合率、F値は向上した。この結果から、SVMと同様にRFで液状化発生予測モデルを構築する際には、地震動の継続時間を説明変数に含めることは有効であることがわかる。

さらにモデルの精度を向上させるため、既往研究³⁶⁾に基づき、液状化の発生傾向別のグループ1~5ごとに液状化発生予測モデルを構築 (RF3) し、グループ分けをしないRF1と比較する。学習データの割合は全体の70%とし、学習データを最も精度良く分類するハイパーパラメータを5分割交差検証法によるグリッドリサーチ

表-6 RF3のハイパーパラメータ

	グループ1	グループ2	グループ3	グループ4	グループ5
mtry	4	3	4	3	5

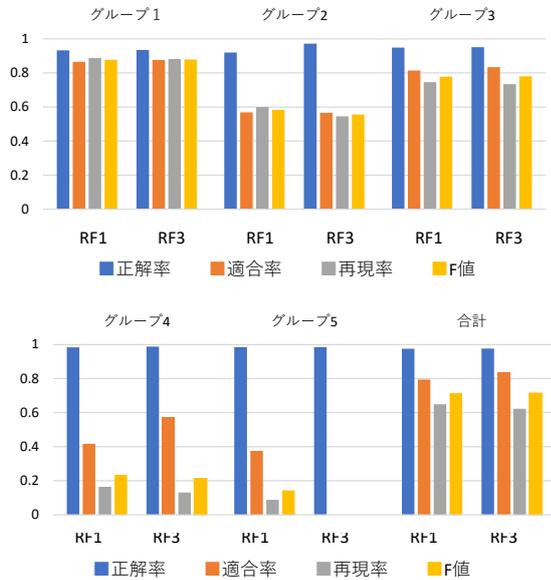


図-6 RF1とRF3の精度評価指標の比較

により定めた(表-6)。残りの30%のテストデータにRF3を適用し精度評価を行った。図-6に、RF1とRF3の精度指標値を比較する。グループごとに見ると、液状化の発生数が多いグループ1、グループ3では、RF1、RF3ともに精度が非常に高く、特にグループ1はF値が0.879、またその他の指標も0.87以上となった。一方、グループ5では、RF3は液状化と予測した地点がなく、再現率は0、適合率とF値は計算できない。そのほかのグループについては、RF1の方が多く液状化を予測し、多少のばらつきはあるものの総じてRF3よりもF値が高くなった。全体としては、RF1の方が液状化の見落としが少ないという結果になった。

テストデータに対する分類結果を図示したものが図-7、図-8である。どちらもSVMと同様に東京湾岸と利根川周辺での液状化が集中している地域では的中率が多いものの、内陸部での見落としが多い。また、RF1は、RF3と比較して、市川市周辺の液状化を多く予測している。

4. まとめ

本研究ではサポートベクターマシンと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、地震動の継続時間を考慮した液状化発生予測モデルの構築を行った。SVMとRFのどちらにおいても、地震動の継続時間を考慮することで精度が向上すること

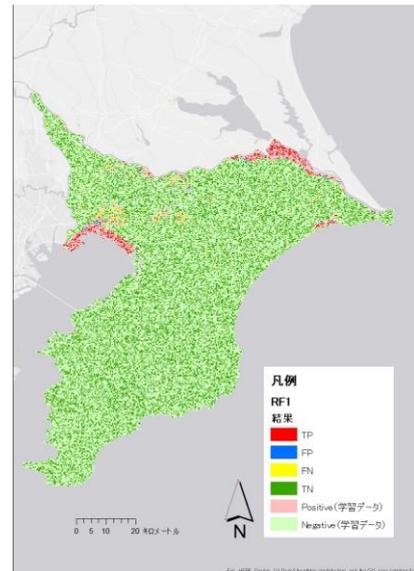


図-7 RF1の予測結果

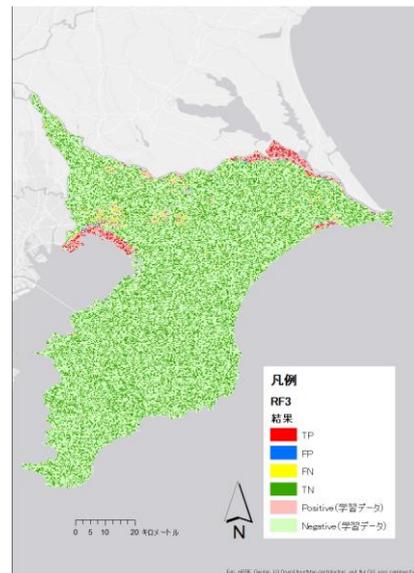


図-8 RF3の予測結果

がわかった。また、液状化の発生傾向を踏まえて微地形をグループ分けした上で、液状化発生予測モデルを構築したが、却って液状化の見落としが増え、精度の向上にはつながらなかった。SVMとRFを比較すると、RFの方が精度が良かった。

今後は、東北地方太平洋沖地震の際の別地域や液状化が発生した他の地震において液状化発生予測モデルを適用し、その汎用性を検証する必要がある。また、過去の地震による液状化履歴の有無、地下水位や地盤改良など、液状化の発生に影響があると考えられる要素を考慮したパラメーターの見直しが必要である。

謝辞: 本研究では、防災科学技術研究所の強震観測網の観測記録を使用した。また、千葉県強震観測網の観測記

録を使用した。記して、謝意を表する。

参考文献

- 1) 若松和寿江, 先名重樹, 小澤京子: 2011年東北地方太平洋沖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. 1, pp. 43-62, 2017.
- 2) 猪股渉, 乗藤雄基, 石田栄介, 塚本博之, 山崎文雄: 東日本大震災における東京ガスの設備被害の概況と超高密度観測情報に基づく低圧ガス導管被害推定の精度検証, 日本地震工学会論文集, Vol. 13, No. 2, pp. 37-44, 2013.
- 3) 岩崎敏夫, 龍岡文夫, 常田賢一, 安田進: 地震時地盤液状化の程度の予測について, 土と基礎, Vol. 28-4, No. 1164, pp. 23-29, 1980.
- 4) 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.
- 5) 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.
- 6) 桑原光平, 松岡昌志: 機械学習を用いた日本全国の液状化危険度の推定, 日本地震工学会論文集, Vol.21, No.2, pp. 2_70-2_89, 2021.
- 7) 柳瀬匡雄, 丸山喜久: 液状化地点の予測に向けた共分散構造分析とサポートベクターマシンの利用, 土木学会論文集 A1 (構造・地震工学), Vol. 75, No. 4, pp. I_133-I_143, 2019.
- 8) 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.
- 9) 先名重樹, 松岡昌志, 若松和寿江, 翠川三郎: 液状化発生率におよぼす強震動の継続時間と地域性の影響, 日本地震工学会論文集, Vol. 18, No. 2, pp. 82-94, 2018.
- 10) 防災科学技術研究所: 地震ハザードステーション, <http://www.j-shis.bosai.go.jp/>
- 11) 防災科学技術研究所: 強震観測網, <https://www.kyoshin.bosai.go.jp/kyoshin/>
- 12) 酒井豊, 楠田隆, 加藤晶子, 宇野亨: 千葉県強震観測網で観測された波形データの公開, 日本地質学会学術大会講演要旨, p. 195, 2000.
- 13) 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/>
- 14) 能島暢呂, 高島拓也: 累積パワーに基づく地震動継続時間の等価振幅レベルに関する考察, 日本地震工学会論文集, Vol. 16, No. 6, pp. 1-15, 2016
- 15) 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.
- 16) 波部斉: ランダムフォレスト, 情報処理学会研究報告, pp. 1-8, 2012.
- 17) 平井有三: はじめてのパターン認識, 森北出版, 2012.

Fundamental Study on Prediction of Liquefaction Occurrence by Machine Learning Considering the Duration of Earthquake Motion

Hiroto NAKAYAMA, Wen LIU and Yoshihisa MARUYAMA

In the 2011 Tohoku Japan Earthquake, liquefaction occurred in various areas of the eastern Japan. In the current liquefaction prediction method, the most common method is calculated using geotechnical survey data, and it gives accurate results. However, it is difficult to apply this method to a wide area. Therefore, it is desirable to establish a method that can easily evaluate the occurrence of liquefaction damage extensively. In this study, we aim to construct a liquefaction prediction model using support vector machine and random forests. Based on the results of previous studies, the duration of seismic motion, which affects the occurrence of liquefaction, is newly added as an explanatory variable. In addition, the models are grouped according to the tendency of liquefaction occurrence based on the Japan Engineering Geomorphologic Classification Map, aiming to develop a model with high prediction accuracy.