上水道管路の地震被害位置の予測に向けた機械学習手法の利用

Use of Machine Learning Techniques to Estimate the Locations of Earthquake Induced Damage to Water Distribution Pipes

18WM1344 有田 京平 Kyohei Arita 指導教員 丸山 喜久

SYNOPSIS

Water distribution pipelines in Sendai City, Miyagi Prefecture were heavily damaged by the 2011 Tohoku Japan earthquake. This study employs machine learning techniques to estimate the locations of damage incidents to water distribution pipelines during an earthquake. The authors performed a series of logistic regression analysis to estimate the damage probability of water distribution pipes. Besides, support vector machine and random forests were employed to perform a binary classification. The results were evaluated based on accuracy indices, such as true positive rate, true negative rate, overall accuracy, Matthews Correlation Coefficient, and so on.

1. はじめに

2011 年東北地方太平洋沖地震では、東北地方から関東地 方にかけて上水道施設に多くの被害が発生し、約 256.7 万 戸が断水した¹⁾. 宮城県仙台市では、上水道施設の被害総 数は 1,064 件であり、導水・送水・配水管の被害数は 437 件 であった²⁾. 表-1 に仙台市の導水・送水・配水管の主な管 種ごとの被害数と管路延長を示す.

表-1 仙台市の導水・送水・配水管の被害数と延長

管種	被害数	延長(km)
DIP:ダクタイル鋳鉄管	117	2723.1
VP:塩化ビニル管	297	1514.5
SP:鋼管	12	137.3
LP:鉛管	2	3.4
PP:ポリエチレン管	3	52.7
GP:亜鉛メッキ管	6	4.9
合計	437	4458.0

地震調査研究推進本部では、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、水道施設への大きな影響も危惧される.このこと から、地震による水道管の被害が発生する危険性がある場 所を予測することは、事前の防災対策の立案などに有意義 であると考えられる.

既往研究 ^{3),4)}では,被害状況の分析や水道管の地震被害 予測式の予測精度に関する検討が行われてきた.水道技術 研究センターは地震による水道管路の被害率*R_m*(件/km) の推定式を以下のように提案している.

$$R_m = C_p C_d C_g R(v) \tag{1}$$

ここで、 C_p は管種・継手補正係数、 C_d は口径補正係数、 C_g は微地形補正係数、vは地震動の地表面最大速度 (PGV) である. R(v)は標準被害率と呼ばれ、式(2)で表される.

R(v) = 9.92×10⁻³×(v-15)^{1.14}(2)式(1)から算出した VP の被害数と、実際の被害数を仙台市



図-1 VPの実際の被害位置(左)と水道技術研究セン ターの被害予測式から算出した被害数(右)

の供給ブロックごとに集計したものを図-1に示す.図-1より,被害数および被害位置の的中には至っていないエリアが多いことがわかる.

そこで、本研究では機械学習手法による被害位置の予測 を行う.そして、予測結果を評価し、今後の地震による上 水道管路の被害の予防・軽減、早期復旧に貢献することを 目指す.

2. 分析対象データ

本研究では、東北地方太平洋沖地震の際の宮城県仙台市 を対象地域とした.使用したデータは導水・送水・配水管 の被害データ、微地形区分、PGV、管路延長、標高、傾斜 角度、敷設年および口径である.上水道管の管種に関して は、管路延長および被害数の割合が大きい DIP(一般継手) と VP に着目して分析を行った.DIP(一般継手)と VP の 敷設状況および被害位置を図-2 に示す.

微地形区分は、水道技術研究センターの管路被害予測式 の微地形補正係数の区分に基づき、カテゴリ変数である微 地形区分をダミー変数化して使用する(表-2).なお、カテ ゴリ数が3以上の変数をダミー変数化して用いる場合は、 ある1つのグループを基準値としなければならないため、 本研究では微地形補正係数が1.0のグループを基準値とした.

PGV は、丸山ら⁵⁾が地形条件を踏まえて定めた地震計の 影響範囲内のデータを使用した.標高および傾斜角度は、



図-2 DIP (一般継手) と VP の導水・送水・配水管の 敷設状況および被害位置

表-2 微地形区分のグループ分け

微地形	補正係数	グループ
山地, 山麓地, 丘陵, 火山地, 火山山麓地, 火山性丘陵	0.4	g_1
砂礫質台地、ローム台地	0.8	g_2
谷底低地,扇状地,後背湿地, 三角州·海岸低地	1.0	g_3
自然堤防, 旧河道, 砂州•砂礫洲, 砂丘	2.5	g_4

国土交通省が整備する国土数値情報から取得した.本研究 で用いた微地形区分のデータは5次メッシュ(約250mメ ッシュ)ごとに整備されているため,解析のメッシュサイ ズを5次メッシュとした.

本研究では、被害位置の予測を管種の違いを考慮せず DIP と VP をまとめて行った.分析対象のメッシュ数を表-3 に示す.本研究で使用したデータは、正例(被害あり) と負例(被害なし)のデータ数に大きな偏りがあり,分類 精度に影響するため、被害ありのメッシュにデータ数の比 に基づいた重みを付けて不均衡データを補正した.また、 機械学習に適用する際、分析対象データの70%を教師デー タ、30%をテストデータとした.

3. 被害位置の予測および評価

本研究では、4 つの機械学習手法により被害位置を予測 した.予測結果の評価方法として正答率、適合率、再現率、 F 値、マシューズ相関係数の5つの評価指標を導入する. 各指標式は式(3)~式(7)の通りであり、式中のTP, FP, FN, TN は表-4 のように定義される. (2)

正答率(Accuracy) =
$$\frac{TP + TN}{TP + FP + FN + TN}$$
 (3)

$$IP + FP + FN + IN$$
 (4)
適合率(Precision) = TP

再現率(Recall) =
$$\frac{TP}{TP + FN}$$
 (5)

$$F \ \ (F - measure) = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$
(6)

$$=\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(FN + TN)(TN + FP)(FP + TP)}}$$
(7)

3.1 ロジスティック回帰分析

ロジスティック回帰分析 %は、一般にある事象の発生す

表-3 分析対象のメッシュ数の内訳

該当 メッシュ数	被害あり メッシュ数	被害なし メッシュ数	重み
1839	212	1627	7.67倍

表-4 予測精度評価のための混同行列

$\overline{}$	_	実際のデータ			
		被害あり True	被害なし False		
予測	被害あり Positive	TP (True Positive)	FP (False Positive)		
結果	果 被害なし Negative	FN (False Negative)	TN (True Negative)		

る確率Pを,その現象の生起を説明するために観測された 量的・質的な複数の変数 $x = (x_1, x_2, ..., x_n)$ で説明しようと考 える場合に, $x = (x_1, x_2, ..., x_n)$ という状態の下で現象が発生 する条件付き確率をP(x)で表す回帰分析である.

ある事象の発生確率P(x)を算出するためのロジスティック回帰モデルを式(8)に示す. β_0 は定数, $\beta_1 \sim \beta_n$ は回帰係数, $x_1 \sim x_n$ は説明変数である.

$$P(x) = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)\}}$$
(8)

ロジスティック関数から算出される被害の発生確率P(x)は0から1の間の値をとる.

ロジスティック回帰分析の結果を ROC (Receiver Operating Characteristics) 曲線を用いて検証する.ここで, 被害ありと正しく判別する率を敏感度 TPR (sensitivity) と し,被害なしと正しく判別する率を特異度 TNR (specificity) とする.敏感度 TPR および特異度 TNR は, 表-4 の混同行 列に基づき,式(9), (10)で表される.

$$TPR = \frac{TP}{TP + FN}$$
(9)

$$TNR = \frac{TN}{TN + FP}$$
(10)

閾値を適宜変化させ TPR, TNR をその都度算出し, 横軸に 1-TNR, 縦軸に TPR をとったものを ROC 曲線という. 判 別モデルが有効である場合, ROC 曲線は 1-TNR, TPR の比 が 1:1 の直線から左上に離れていく. ROC 曲線下の面積は AUC (Area Under the Curve) と呼ばれ, 最大値は 1, ラン ダムな判別では 0.5 となる.

ロジスティック回帰分析の予測結果と実際のデータを比較したものを表-6(a)に示す.また,表-5 および図-3 にロジスティック回帰モデル式の各項と説明変数の関係とロジットモデルの回帰係数を示す.仙台市における東北地方太平洋沖地震の上水道埋設管の被害が丘陵地の宅地造成地盤で被害が集中していることが示されているように^{5),7)},丘陵を含む微地形区分のグループの説明変数の回帰係数の値より大きいことがわかる.また,VPの管路延長も比較的被害発生への影響が大きい.

3.2 ランダムフォレスト(RF)

ランダムフォレスト(RF)⁸⁰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(11)に示す情報利得 IG が最 大となるようにする.



図-3 ロジスティック回帰分析 モデル式の回帰係数

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I(D_j)$$
(11)

ここで、D_nは教師データ、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(12)はエントロピーの式を表している.

$$I(N) = -\sum p(k|N)logp(k|N)$$
(12)

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,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. R 言語を用いたランダムフォレス トでは,この mtry をハイパーパラメータとして分類器を構 成する.

RF の予測結果と実際のデータを比較したものを表-6(b) に示す. また, RFによる説明変数の影響度を表-5 および 図-4 に示す. 標高と VP に関する説明変数の影響が大きい 結果となった.

3.3 サポートベクターマシン(SVM)

サポートベクターマシン (SVM)⁹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である.本研究では, Gaussianカーネルを用いたSVMを行う.

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、 ϕ は入 カベクトルxを特徴空間Fへ非線形写像する関数(式(14))、 ζはスラック変数、bはスカラー変数を表す.

$$\min_{w,b,\zeta} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \zeta_i$$
(13)

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

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

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.

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



図-4 RF 説明変数の影響度

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ため、グリッドサーチに基づき適切なγとCを設定する.γ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γとCの組み合わせを求める.本研究では、γの範囲を10⁻⁵~10⁵、Cを10⁻²~10²に設定して、10分割の交差検証法によって適切なパラメータを求めた.グリッドサーチにより算出したパラメータを用いて分析を行ったところ、すべてのテストデータが被害なしと分類された.これには過学習の可能性があったため、デフォルトのパラメータ(γ:特徴ベクトルの次元数の逆数、C=1)でも分析した.

ータを比較したものを表-6(c)に示す.再現率は SVM が最 も高く,被害発生地点の検出率が高かった.図-5 に SVM の被害発生地点評価図を示す.

3.4 1 クラス SVM

1クラスサポートベクターマシン(OCSVM)¹⁰は,SVM をデータの高密度領域を推定する領域判別問題に適用した 教師なし学習のアルゴリズムの異常検知手法である.すべ ての例を特徴空間に写像して,原点と大部分の例を最大マ ージンで分離することにより,異常例の集合を検知する.

l個の学習データ $\mathbf{x}_1, \dots, \mathbf{x}_l$ を与える.ただし, \mathbf{x}_i は空間 \mathcal{R}^d における点である. ϕ をデータ空間 \mathcal{R}^d の点を高次元特徴空間へと写像する関数とする.xが正常例か異常例であるかを決定する決定関数 $f(\mathbf{x})$ は,次式で与えられる. $f(\mathbf{x})$ の符号が負の場合 \mathbf{x} は異常例であり,それ以外の場合は正常例となる.

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho \tag{16}$$

ここで、wは法線ベクトルであり、wとρにより境界面が決 まる.以下の最適化問題を解くことで、大部分の例と原点 とを最大マージンで分離することができる.

$$\min_{w,\xi,\rho} \frac{1}{2} \|w\|^2 + \frac{1}{vm} \sum_i \xi_i - \rho$$
(17)

s.t. $\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle \ge \rho - \xi_i, \xi_i \ge 0$

ここで、パラメータ $v \in (0,1)$ は、全学習例の内の異常値の割 合の上限であり、全例の内のサポートベクトルの割合の下 限である. ξ_i はスラック変数であり、式(17)の不等式に示す ように最適化の制約を緩和する.式(17)をラグランジュの 未定乗数法で解くことで、パラメータ w,ξ,ρ を決定する.こ の決定方法は、2クラス分類学習用の SVM と類似する.



図-5 SVM の被害発生地点評価図

本研究では、vを被害ありと被害なしのデータ数の比か ら 0.13 とし、ξを 0.1~2.0 の範囲で調整した. OCSVM の 予測結果と実際のデータを比較したものを表-6(d)に示す. MCC は 0.083 であり,他の手法に劣る結果となった.また, 正答率は良好であったが,適合率・再現率は低かった.

4. 予測結果の多数決による評価

本研究で用いた4手法のうち,予測精度が比較的高かっ たロジスティック回帰分析・RF・SVMの3手法で予測結 果の多数決を取り,各手法の精度と比較した.多数決の結 果のうち3手法とも同じ予測結果であったもの(表-7)に 関しては,再現率以外の指標が各手法の指標よりも向上し た.特に,3手法とも被害なしと予測したデータの約96% が的中し,実際に被害が発生していない確率が高かった.

表−7	3	手法と	も同	じ予測結果のデー	タ
-----	---	-----	----	----------	---

				実際のデータ				
			_	被害な	あり	被	害なし	
	予測	被害あり		22		43		
	結果	被害なし	,	15		327		
[正答率	適合率	Ŧ	再現率	F	直	мсс	
Γ	85.7%	33.8%		59.5%	0.4	31	0.375	

5. まとめ

本研究では、ロジスティック回帰分析, RF, SVM, OCSVM の4つの機械学習手法を用いて、東北地方太平洋沖地震の 際の宮城県仙台市の上水道管路の被害のうち、管路延長・ 被害数ともに大部分を占める DIP と VP に着目し、被害発 生地点の予測を行った。

MCCは,教師あり機械学習手法のロジスティック回帰分

表-6 各手法の分析結果

(b)	RF	(mtry=2)
---	----	----	----------

				実際のデータ			
			被害あり		被害なし		
	予測 被害あり)	26		56	
	結果	結果被害なし		38		432	
1	E答率	適合率	Ē	耳現率	F	直	MCC
	83.0%	31.7%	4	40.6%	0.3	56	0.262

(d) $1 \not \neg \neg \neg X$ SVM ($v = 0.13, \xi = 0.20$)

				実際のデータ				
				被害あり		被	害なし	
	予測	 予測 被害あり 結果 被害なし 		19	19		94	
	結果			45		394		
[]	正答率	適合率	Ŧ	再現率	Fĺ	直	MCC	
	74.8%	16.8%	1	29.7%	0.2	15	0.083	

析, RF, SVM の 3 つの手法ではそれぞれ 0.244, 0.262, 0.274 で,良好な結果が得られた.一方,適合率はロジステ ィック回帰分析が 26.7%, RF が 31.7%, SVM が 23.7%と 低く,過剰に被害ありと予測してしまうケースも多く見ら れたため,説明変数の選択方法や学習データの見直しを行 う必要があると考える.被害の発生に対する説明変数の影 響度合いは,ロジスティック回帰分析の結果からは丘陵を 含む微地形区分のグループと VP の管路延長が大きく, RF の結果からは標高と VP に関する情報が大きいことが示さ れた.また,ロジスティック回帰分析・RF・SVM の 3 手法 とも被害なしと予測したデータうちの約 96%が的中し,実 際に被害が発生していない確率が高いことがわかった.

今後は、予測精度の向上に向けて被害形態ごとに分けた 分析や管種ごとの被害の特徴の見直しを行い、管種ごとに 被害予測の際に用いる説明変数を変えるなど、より詳細な 分類モデルの構築を目指す.また、他の機械学習手法に関 する検討や、予測モデルの汎用性の検討を行う.

参考文献

- 厚生労働省健康局水道課:東日本大震災指導施設被害 状況調査最終報告書,2013.
- 仙台市水道局:東日本大震災 仙台市水道復旧の記録, 2012.
- 3) 水道技術研究センター:地震による管路被害予測の確立 に向けた研究報告書, 2013.
- 永田茂,丸山喜久,庄司学:既往の上水道配水管の地震 被害予測式の予測精度に関する検討,土木学会論文集 A1 (構造・地震工学), Vol.71, No.4, pp.50-60, 2015.
- 5) 丸山喜久,伊藤智大,若松加寿江,永田茂:東北地方太 平洋沖地震における宮城県仙台市の水道管路被害状況の 分析,土木学会論文集 A1(構造・地震工学), Vol.70, No.3, 377-388, 2014.
- 6) 青木繁申:Rによる統計解析,オーム社,2009.
- 永田茂,西野雅夫,鈴木清一:東日本大震災における上 水道管路施設の被害分析,土木学会 67 回年次学術講演 会概要集,pp.415-416, 2012.
- 波部斉:ランダムフォレスト,情報処理学会研究報告, 2012.
- 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.
- 10) 藤田裕文,松川徹,鈴木英之進:マルチタスク学習用 1 クラス SVM を用いた新規人物に対する特異顔表情検知, 情報処理学会研究報告, 2017.